Evaluation of the DC Opportunity Scholarship Program

Impacts Three Years After Students Applied

Technical Appendix

Ann Webber, Westat

Ning Rui, Westat

Roberta Garrison-Mogren, Westat

Robert B. Olsen, Westat

Babette Gutmann, Westat

Meredith Bachman, Project Officer, Institute of Education Sciences

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

NCEE 2019-4006
U.S. Department of Education

Evaluation of the DC Opportunity Scholarship Program

Impacts Three Years After Students Applied

Technical Appendix

Ann Webber
Ning Rui
Roberta Garrison-Mogren
Robert B. Olsen
Babette Gutmann
Westat

Meredith Bachman
Project Officer
Institute of Education Sciences

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

NCEE 2019-4006
U.S. Department of Education

U.S. Department of Education
Betsy DeVos
Secretary

Institute of Education Sciences
Mark Schneider
Director

May 2019

This report was prepared for the Institute of Education Sciences under Contract No. ED-IES-12-0018. The project officer was Meredith Bachman in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

To order copies of this report,

- Write to ED Pubs, U.S. Department of Education, P.O. Box 22207, Alexandria, VA 22304.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 1-877-576-7734.
- Fax your request to 703-605-6794.
- Order online at www.edpubs.gov.

This report also is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

Contents

Appendix Tables

Appendix Figures

This page intentionally blank.

Appendix A. Program Features, Lottery Structure, Study Sample, and School Characteristics

This appendix describes key features of the Opportunity Scholarship Program (OSP), how the lottery for scholarships was conducted, characteristics of the student sample, and characteristics of the schools that eligible scholarship applicants attended.

A-1. Program Features

The Scholarships for Opportunity and Results (SOAR) Act requires the OSP to be operated through a federal grant to a local entity, and to be supervised by the U.S. Department of Education's (the Department's) Office of Innovation and Improvement, and the Office of the Mayor of the District of Columbia (DC). In August 2015, the Department awarded a three-year grant to a DC-based nonprofit organization, Serving Our Children, to implement the OSP. Another nonprofit, the DC Children and Youth Investment Trust, administered the OSP between 2011 and August 2015.

The program operator is responsible for ensuring that participating schools meet reporting requirements and financial responsibilities. Schools must provide accreditation information, ensure that teachers in core subjects have a baccalaureate degree or higher, and assure compliance with the statute's language prohibiting discrimination against applicants on the basis of race, color, national origin, religion, or sex. Schools also have to have financial systems and procedures, and submit proof of adequate financial resources if the school has been operating for five years or less. The operator of the program also is responsible for setting up the application process, recruiting applicants and schools, awarding scholarships, and monitoring awardees and schools. The SOAR Act does not specify that monitoring should take into account the academic performance of participating private schools or of OSP students in the schools.

Families apply for the scholarship and the program operator determines their eligibility (see exhibit 1 in chapter 1). Eligible families who receive scholarship offers then decide which participating private schools—if any—they will apply to, and those schools decide if applying families meet their admissions criteria, which schools set on their own. The legislation expressly states that participating schools do not have to alter or change their tuition or their admission criteria for OSP scholarship students. Students can be offered a scholarship but not be admitted to a private school they want to attend. There is no obligation to use the scholarship. Eligible families who do not receive scholarship offers also can apply for and attend participating private schools, but receive no scholarship support.

A-2. Lottery Structure

The evaluation includes three consecutive cohorts of students from lotteries conducted in 2012, 2013, and 2014 (in late spring or early summer of each year).[1] A total of 1,771 students applied for and were eligible to enter the lottery for scholarships in these three years. The OSP program operator conducted the annual lotteries using a computer program designed by the study team, with the execution of the lotteries supported by the study team and observed by staff from the Department.

The OSP statute specifies a higher probability of award for applicants in three priority groups: (1) siblings of students already participating in the program, (2) students attending a low-performing school designated as in need of improvement (SINI) at the time of application, and (3) students previously offered a scholarship who did not use it. The relative probabilities for each group were determined as follows by the Department officials who oversaw the program:

- Twenty-five percent higher probability for SINI and previous awardees who never used a scholarship, and
- Forty percent higher probability for applicants with a sibling already in the OSP.

The probabilities were stated in percentage terms and were applied relative to the probability for the "no priority" group. Because the number of eligible applicants in each group differed each year of the lottery, the absolute or actual award probability for each priority group differed somewhat but the relative priorities stayed the same across the years (table A-1).

Table A-1. Scholarship offers by priority group categories, application year and treatment status

Application year and treatment status	Total	No priority	Sibling already in program	Attended SINI school or previous awardee never used
2012				
Treatment	316	46	47	223
Control	220	49	23	148
Award probability	59%	48%	67%	60%
2013				
Treatment	394	87	62	245
Control	324	103	36	185
Award probability	55%	46%	64%	57%
2014				
Treatment	285	84	44	157
Control	232	95	24	113
Award probability	55%	47%	65%	58%

NOTE: Students in more than one category (i.e., a sibling already in the program *and* enrolled in SINI school) were given the probability for the higher of the two categories.
SOURCE: OSP applications and records from OSP program operator.

[1] A lottery was not conducted in 2011, the first year after the OSP was reauthorized. That year, all eligible applicants were offered a scholarship, and therefore, that cohort of applicants could not be used in this experimental evaluation.

The lotteries yielded scholarship offers to 995 students, 56 percent of eligible applicants (table A-2). Because the lotteries (essentially a flip of a coin) determined which students were in the treatment and control groups, the two groups were expected to have similar characteristics—ones that could be observed, such as age, gender, and income, as well as ones that could not be observed or were difficult to observe, such as motivation to succeed in school and desire to attend a private school.

Table A-2. OSP scholarship offers by study cohort

Study cohort (year of application)	Number of eligible applicants (full sample)	Scholarship offered (treatment group)		Scholarship not offered (control group)	
		Number	Percent	Number	Percent
2012	536	316	59	220	41
2013	718	394	55	324	45
2014	517	285	55	232	45
Total	1,771	995	56	776	44

SOURCE: OSP applications.

A-3. Characteristics of the Study Sample

Families applying for a scholarship completed an application that included information about various student and family characteristics. At the time of application, students also completed the *TerraNova* reading and mathematics tests. Table A-3 shows these characteristics for the full sample of eligible applicants. The observed differences between characteristics of the treatment and control groups, at the time of application, mostly arose from sampling variation. Differences were statistically significant for only one of the 29 characteristics, how long students had been living at their current address, which was 69 months for students in the treatment group and 62 months for students in the control group.

Average test scores at the time of application were similar for treatment and control group students within grade bands (i.e., students entering grades K–2, 3–5, 6–8, and 9–12 when they applied). The *TerraNova* is vertically scaled and its average scores are higher in higher grades, which is consistent with the average baseline reading scores of 481 for grades K–2 and 670 in grades 9–12.

Table A-3. Characteristics of treatment and control groups at time of application (full sample)

Characteristic	Treatment			Control			Difference of means
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	
Year of application							
First cohort (spring 2012)	995	30.0%	45.8	776	30.0%	45.8	0.0
Second cohort (spring 2013)	995	41.0	49.0	776	41.0	49.0	0.0
Third cohort (spring 2014)	995	29.0	45.0	776	29.0	45.0	0.0
Entering grade							
Kindergarten	995	23.0%	42.1	776	27.0%	44.4	4.0
Grade 1	995	12.0	32.0	776	10.0	31.0	2.0
Grade 2	995	9.0	29.0	776	10.0	30.0	-1.0
Grade 3	995	10.0	30.0	776	8.0	28.0	2.0
Grade 4	995	8.0	27.0	776	8.0	28.0	0.0
Grade 5	995	6.0	24.0	776	5.0	23.0	1.0
Grade 6	995	9.0	29.0	776	7.0	26.0	2.0
Grade 7	995	6.0	24.0	776	6.0	23.0	0.0
Grade 8	995	4.0	20.0	776	5.0	22.0	-1.0
Grade 9	995	6.0	23.0	776	8.0	27.0	-2.0
Grade 10	995	4.0	18.0	776	4.0	19.0	0.0
Grade 11 or 12[1]	995	3.0	16.0	776	3.0	16.0	0.0
Test score							
Reading scale score at time of application	968	561.0	91.3	747	562.5	94.7	-1.5
Grades K–2	422	480.9	55.1	347	481.4	66.8	-0.5
Grades 3–5	236	595.3	48.3	166	595.8	60.8	-0.4
Grades 6–8	193	637.4	40.9	132	639.6	46.6	-2.2
Grades 9–12	117	669.8	34.7	102	670.2	40.2	-0.4
Mathematics scale score at time of application	951	534.8	113.5	726	540.8	113.2	-6.0
Grades K–2	406	436.4	67.1	326	441.2	71.0	-4.8
Grades 3–5	235	565.9	60.4	166	570.0	71.8	-4.1
Grades 6–8	193	627.4	54.3	132	631.7	64.3	-4.3
Grades 9–12	117	680.0	50.3	102	677.4	58.4	2.6
Student characteristics							
Student is female	995	49.0%	50.0	776	49.0%	50.0	0.0
Student is African American	995	84.0%	36.0	776	87.0%	34.0	-3.0
Student has disabilities or other challenges	995	15.0%	35.0	776	13.0%	33.0	2.0
Student attends a school in need of improvement	995	64.0%	48.0	776	63.0%	48.0	1.0
Student age difference from median age of grade	995	<0.1	0.5	776	<0.1	0.5	<0.1

See notes at end of table.

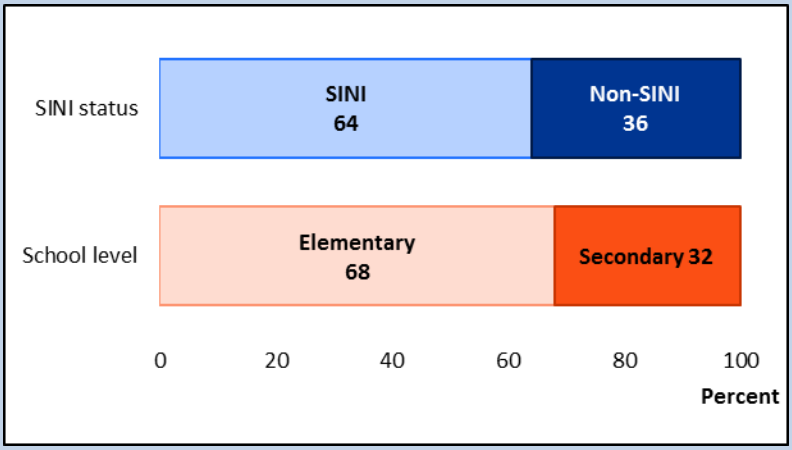
Table A-3. Characteristics of treatment and control groups at time of application (full sample)—Continued

Characteristic	Treatment			Control			Difference
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	
Family characteristics							
Parent went to college	991	60.0%	49.0	768	59.0%	49.0	1.0
Parent gave school grade of A or B at time of application	870	59.0%	49.0	691	57.0%	50.0	2.0
Parent perception of school safety at time of application	890	74.0%	44.0	703	70.0%	46.0	4.0
Parent was employed at time of application	991	48.0%	50.0	769	47.0%	50.0	1.0
Family income in thousands at time of application	995	12.6	13.4	776	13.0	13.5	-0.4
Number of children in household at time of application	984	2.6	1.4	769	2.6	1.4	-0.1
Months at current address at time of application (in tens)	981	6.9	8.5	767	6.2	7.3	0.8*

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.
[1]The percentages for grades 11 and 12 were combined due to small sample sizes.
NOTE: The sample was weighted by the inverse of the probability of being selected in the lottery. For binary variables (e.g., grade level or female), the mean was the proportion of positive responses, and the standard deviation measured how spread out the distribution was from that proportion.
SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered at the time of application.

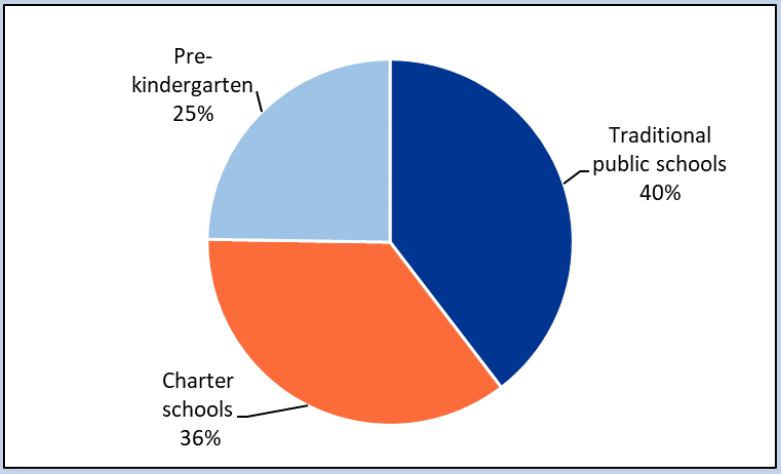
The figures below show selected characteristics for the full sample of eligible applicants at the time they applied for the scholarship. These characteristics include the percentage of students who attended a SINI school and elementary/secondary grades (figure A-1) and the type of school they attended (traditional public, charter, or private; figure A-2). Figures A-3 and A-4 show the grade level the student was entering and the expected grade level for the full student sample in the third year after applying.

Figure A-1. Percentage of eligible applicants, by SINI status and school grade level at time of application



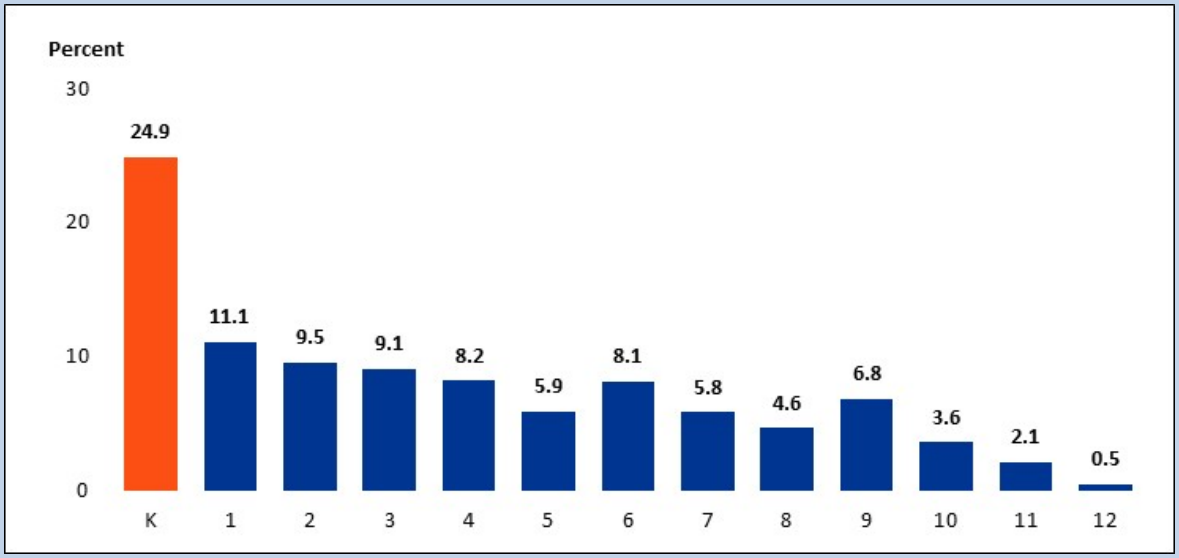
SOURCE: OSP applications.

Figure A-2. Percentage of eligible applicants, by school type at time of application



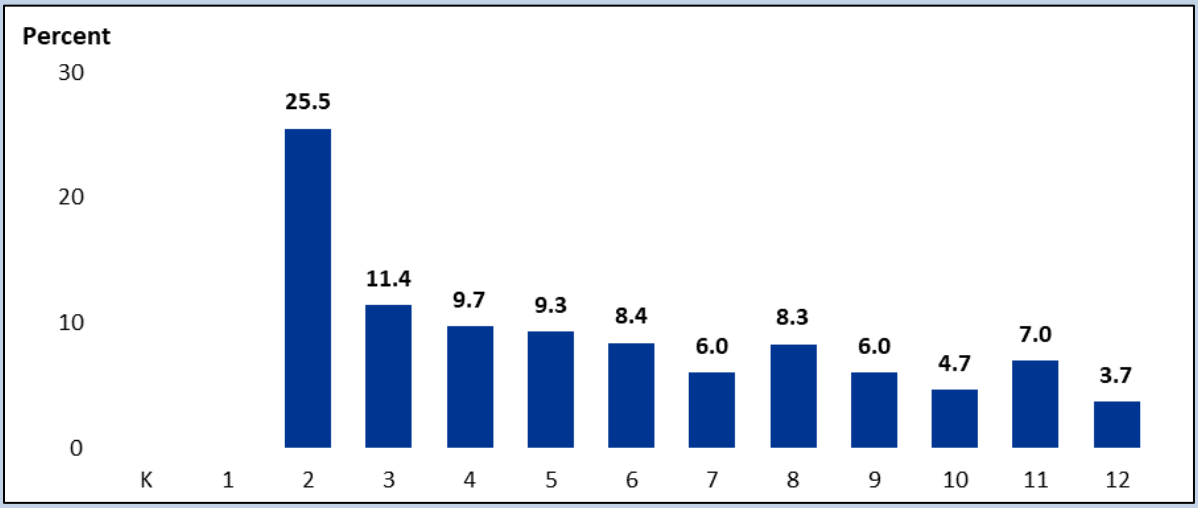
SOURCE: OSP applications.

Figure A-3. Percentage of eligible applicants, by entering grade level at time of application



NOTE: Percents may not sum to 100 because of rounding.
SOURCE: OSP applications.

Figure A-4. Percentage of eligible applicants, by expected grade level three years after application

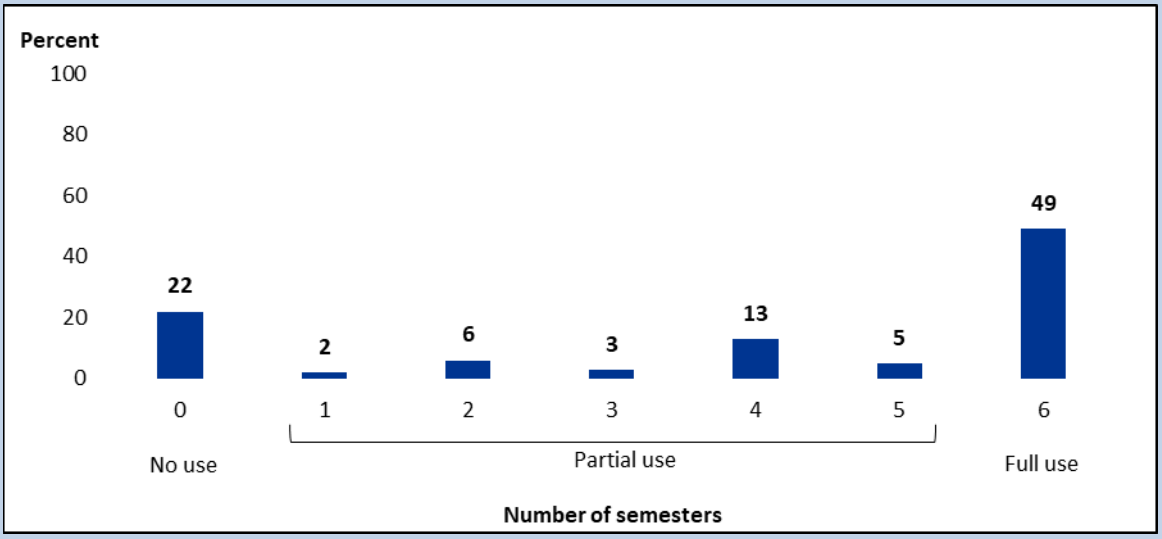


NOTE: Percents may not sum to 100 because of rounding. The expected grade level is two grades above the student's entering grade at the time of application and does not account for students who may have been retained in grade. Students entering grade 11 or 12 at the time of application were no longer part of the study's data collection in the third year and are not shown in this figure.
SOURCE: OSP applications.

Student Participation in the Program

Students who received an *offer* of a scholarship (the treatment group) could decline to use it at all, use it intermittently, say, for one or two semesters, or use it fully. For this report, examining the impacts three years after students and families applied to the OSP, "full use" was defined as using a scholarship for all six semesters, "partial use" as some of the six semesters, and "no use" as none of the semesters. Because the extent of participation was most relevant for understanding program impacts, the participation rates reported here are for the sample of students in the third-year impact sample. This is the group of students who completed a reading achievement test in the third year of followup after applying for a scholarship. Among the third-year impact sample of treatment group students, 49 percent were full users, 29 percent were partial users, and 22 percent did not use it at all (see figure A-5).

Figure A-5. Percentage of treatment group students in the third-year impact sample using the scholarship, by number of semesters of use



SOURCE: Scholarship payment files from Serving Our Children.

Among those students who were offered the scholarship, the rates of scholarship use declined over time (table A-4). This was true for students in the third-year sample and students in the full sample of eligible applicants.

Table A-4. Percentage of treatment group students using the OSP scholarship in each year after application

Year after application	Percent of treatment students using a scholarship	
	Third-year impact sample	Full sample of eligible applicants
Year 1	74	70
Year 2	69	60
Year 3	62	51

NOTE: Sample size was 571 students for the third-year impact sample and 968 students for the full sample of eligible applicants.
SOURCE: Scholarship payment files from Serving Our Children.

A-4. School Characteristics

The kinds of schools that participate in the OSP and that students attend—including those offered a scholarship (the treatment group) and those not offered a scholarship (the control group)—may influence the impact of the OSP. Table A-5 identifies the type of school that students were attending in the spring of the third year after applying for a scholarship. Three years after applying for the scholarship, most students in the treatment group were attending a private school (62 percent), while the others were evenly split between traditional public and charter schools (19 percent in each type). Students in the control group were most likely to be attending a charter school (45 percent) or traditional public school (44 percent), but 11 percent were attending a private school that was participating in the OSP.

Table A-5. Percentage of study participants in the third-year impact sample, by school type, three years after application

	Percent of students	
School type	Treatment group	Control group
Traditional public	19	44
Charter	19	45
Participating private	62	11

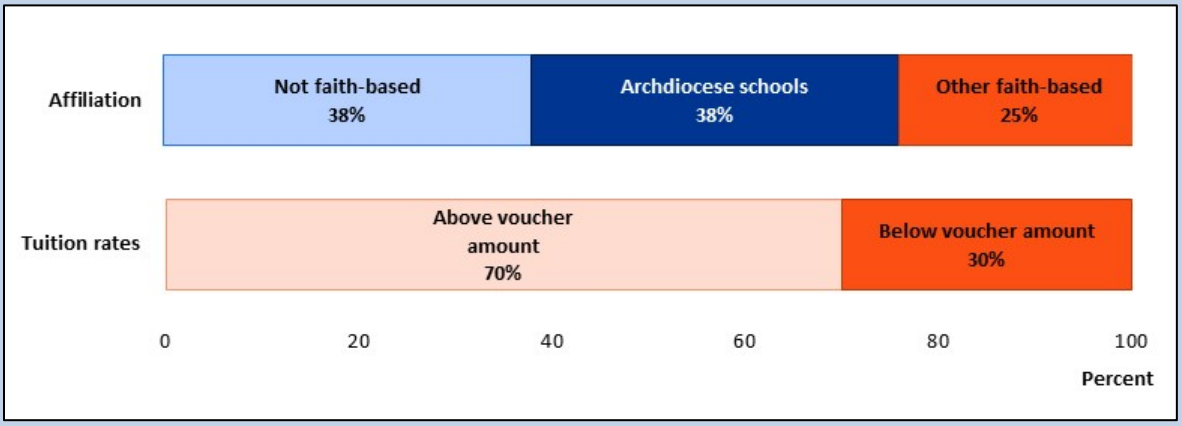
NOTE: The sample was weighted by the inverse of the probability of being selected in the lottery.
SOURCE: School type was obtained at followup testing for students in the third-year impact sample.

Private Schools Participating in the OSP

Private schools participating in the OSP can play a role in the effectiveness of the program, though where students who are offered a scholarship ultimately enroll depends on their families' preferences and the private schools' admissions criteria. The number of private schools participating in the OSP declined from 52 (in the 2013–14 school year) to 49 (in the 2015–16 school year).[2] Of the schools that participated in the OSP in any school year from 2013–14 to 2015–16, 62 percent were religiously affiliated, and 38 percent were Catholic schools operating within the Archdiocese of Washington (figure A-6). Among participating schools, 70 percent had published tuition rates above the maximum voucher amount.[3]

[2] This was a net change. A small number of schools began participating, stopped participating, or closed during this time period.
[3] Among schools where the published tuition rates exceeded scholarship amounts, the average difference was $13,310 (ranging from $177 to $31,519). Tuition amounts used here are ones posted by schools, which can offer other kinds of aid to defray tuition costs. The study's data do not include how much tuition OSP participants actually paid.

Figure A-6. Percentage of participating private schools, by religious affiliation and tuition rates

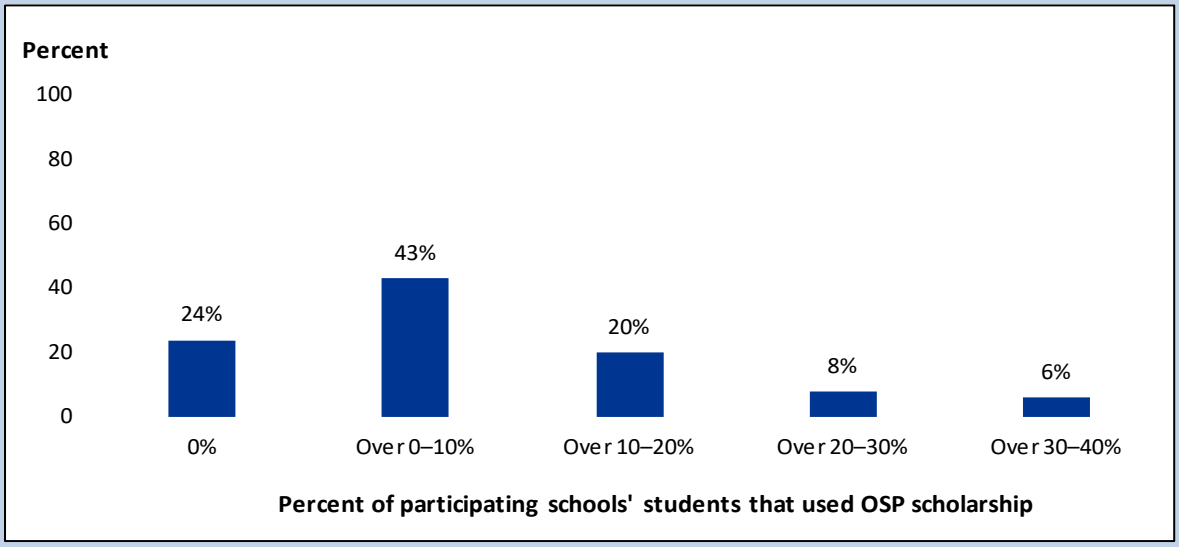


NOTE: Percents may not add to 100 because of rounding. Information presented reflects the 53 private schools that participated in OSP during 2013–14, 2014–15, or 2015–16.

SOURCE: Religious affiliation is from the NCES Private School Survey, 2013–14. Information about tuition rates for OSP participating schools was obtained from the Participating School Directory, published in 2015–16 by Serving Our Children, and in 2013–14 and 2014–15 by DC Children and Youth Investment Trust Corporation.

The proportion of voucher students in participating private schools provides a sense of the extent to which these schools rely on vouchers.[4] On average, OSP students represented 8 percent of enrollment in participating private schools, but the proportion varied widely between schools. During the 2013–14 school year, in 24 percent of participating private schools, there were no OSP students at all, and in 14 percent of participating schools, OSP students represented 21-40 percent of total enrollment (figure A-7).

Figure A-7. Percentage of participating private schools, by the share of OSP students enrolled in their school



SOURCE: NCES Private School Survey, 2013–14 (or 2011–12 or school website); scholarship payment files from Serving Our Children.

[4] An alternate approach would be to analyze the share of revenue private schools received from vouchers, which Hungerman et al. (2017) did for Milwaukee private schools. However, that study relied on data that were not available to this study.

Characteristics of Schools Attended by Students in Treatment and Control Groups

Data from surveys of school principals provide more insight from the school level about differences that the treatment and control group students experienced (table A-6).[5] Compared with students in the control group, students in the treatment group attended schools where principals reported:

- Lower enrollment and lower pupil-staff ratios. For example, school enrollment averaged 289.0 for students in the treatment group and 401.3 for students in the control group.
- Lower use of some school safety measures. For example, 44.3 percent of schools that students in the treatment attended, reported daily presence of police or security staff, compared with 75.0 percent of schools that students in the control group attended.
- Lower suspension rate (8.0 percent compared with 10.7 percent).
- Fewer hours per week of school time (1.7 hours less) and less instructional time in reading and mathematics (about 50 minutes less in reading and 40 minutes less in mathematics per week).
- More frequent tests given by reading and mathematics teachers. For example, among schools that students in the treatment group attended, 89.2 percent of principals reported that testing in mathematics occurred weekly or more often, compared with 77.1 percent of principals at schools that students in the control group attended.
- More availability of instructional programs for advanced learners or talented/gifted students (54.1 percent offered, compared with 40.4 percent) and more availability of individual tutors in school (70.5 percent offered, compared with 65.4 percent).
- Less availability of instructional programs for students with learning disabilities (69.7 percent compared with 89.4 percent) and students learning English (50.4 percent compared with 68.4 percent).
- More availability of differentiated instruction (81.3 percent of schools offered, compared with 79.0 percent).

These average differences in school characteristics are an indication that school environments and instructional experiences differed for the two student groups.

[5] The study administered principal surveys to all schools in DC to collect comparable data for public and private schools. Note that these estimates were affected by the number of students in the study who attended a school. If many students in the study attended large private schools, average enrollment in table A-10 would be larger than average enrollment in all participating private schools. Similarly, if many students in the control group attended large public schools, average enrollment in schools that these students attended would be larger than average enrollment in DC public schools.

Table A-6. Characteristics of schools that students in the third-year impact sample attended, three years after application

Characteristic	Treatment group average	Control group average
Enrollment	289.0	401.3*
Percent African American	74.7%	73.4%
Percent Hispanic	13.7%	17.2%*
Pupil-staff ratio	10.4	11.8*
Safety measures		
Process for screening students using metal detectors	19.0	26.3*
All or most of the students are required to stay on school grounds during lunch	97.3	97.9
Drug sweeps	5.6	4.5
Daily presence of police or security persons	44.3	75.0*
Video surveillance	83.3	91.2*
Mean suspension rate	8.0%	10.7%*
Weekly instructional time (in hours)		
Length of typical school week	30.8	32.5*
Time in mathematics instruction	5.2	5.8*
Time in reading instruction	6.1	6.9*
Frequency of testing English, reading, or language arts skills of students†		
More than once a week	21.3%	15.2%
Weekly	64.6	56.8
Monthly or less often	14.1	28.0
Frequency of testing arithmetic or mathematics skills of students†		
More than once a week	18.5%	22.7%
Weekly	70.7	54.4
Monthly or less often	10.8	22.9
Availability of instructional programs for		
Advanced learners or talented/gifted students	54.1%	40.4%*
Students with learning disabilities	69.7	89.4*
Non-English speakers	50.4	68.4*
Individual tutors available to students in school	70.5%	65.4%*
Differentiated instruction†		
School offers differentiated courses in core curriculum but students have open access to any course provided they have taken the required prerequisite(s)	20.7%	21.4%
School offers differentiated courses and does differentiated grouping in core curriculum	60.6	57.6
School offers a variety of undifferentiated courses in core curriculum and students have open access to any course provided they have taken the required prerequisite(s)	18.7	21.0

* Difference between the treatment group and the control group was statistically significant at the 0.05 level.
†Tests for statistical significance were conducted using a chi-square test and the difference between groups was statistically significant at the 0.05 level.
NOTE: The number of schools providing data for this table varied by characteristic, ranging from 175 to 229 schools. For the treatment group, the number of schools ranged from 142 to 188, and for the control group it ranged from 144 to 185 schools. Because some schools enrolled students from both the control and treatment groups, they contributed to the school characteristics for both groups. School characteristics were weighted by the proportion of students in the study sample attending. Each student was assigned characteristics of their school in the relevant year.
SOURCE: Data for average enrollment, pupil-staff ratio, and race/ethnicity were from the NCES Private School Survey, 2015–16 (for private schools) and from the Common Core of Data, 2015–16 (for public schools). Data for safety measures, suspensions, frequency of testing, instructional programs, tutoring, and differentiation were from the study's principal survey, three years after application. Characteristics for private schools may differ from those previously reported because some participating private schools did not enroll any OSP students, which gave them a weight of zero for these characteristics.

About half (49 percent) of all public school principals reported they were aware of the OSP, while only one-quarter (23 percent) reported making changes to encourage students to remain enrolled in their school (table A-7). Among principals who reported making changes, the most common response to the OSP was adding a parent orientation or meeting to describe the school's offering and performance (86 percent) (table A-8).

Table A-7. Percentage of public school principals reporting awareness of the OSP and changes in response to the OSP

Principal awareness	All public schools	Traditional public schools	Charter schools
Has heard of the DC Opportunity Scholarship Program	49	54	43
Made changes specifically to encourage students to remain enrolled in their school	23	27	19

NOTE: Sample size was 218 principals (112 traditional public and 106 charter schools).
SOURCE: Principal survey administered by the study, spring 2015.

Table A-8. Percentage of public school principals reporting specific changes in response to the OSP

Changes reported	All public schools	Traditional public schools	Charter schools
Added parent orientation or meeting to describe school offerings and performance	86	90	79
Participated in one or more school fairs	69	73	63
Made efforts to improve the physical appearance of your school	67	70	63
Promoted your school through the use of flyers, radio ads, newspapers ads, or other methods of advertising	65	67	63
Offered additional courses (e.g., introduced a course in computer technology or art)	63	63	63
Added tutoring or other special services to help improve academic achievement	61	70	47
Increased school safety provisions	47	53	37
Adjusted disciplinary rules	39	40	37
Altered class sizes	39	40	37

NOTE: Sample size was 49 principals (30 traditional public and 19 charter schools).
SOURCE: Principal survey administered by the study, spring 2015.

This page intentionally blank.

Appendix B. Technical Approach

This appendix provides more detail about aspects of the evaluation that follow from its experimental design, including the study's ability to measure impacts that may be present (statistical power), and the statistical approach to measuring impacts. In addition, it provides technical details about the calculation of percentile changes, outcome measures and data collection procedures, and the construction of sampling and nonresponse weights.

B-1. Measuring the Impact of a Scholarship Offer and Its Use

The lottery created an experiment, with a treatment and a control group that are statistically similar except for the offer of a scholarship, which is a powerful tool for measuring whether the OSP program caused student outcomes to change. The study compares outcomes for the two groups to measure the impacts of a scholarship offer. However, students in the treatment group who *use* their scholarship do not have direct counterparts in the control group—the study did not know which students in the control group would have used their scholarship if it had been offered to them. To measure impacts of use required the study to adjust impacts measured for the full sample. The adjustment procedure is described below.

An implication of the single-lottery structure was that students chose a school *after* the lottery. The study cannot know which schools students in the control group would have chosen had they been offered a scholarship. Researchers have not created ways to adjust impacts that would allow the study to estimate relationships between school characteristics and overall impacts, as they have with the relationship between the offer of a scholarship and its use. As a result, while overall impacts of the OSP are measured rigorously, sources of impacts cannot be measured at that level of rigor.

B-2. Detecting Impacts

The term *power* refers to a study's ability to detect impacts, which means to find that impacts are statistically significant when they in fact arise. Finding that an impact is statistically significant when it does not arise also is possible and is controlled in statistical tests by setting a Type I error rate in statistical tests.

A study's power is related to its sample size and statistical properties of outcomes being measured. For the same outcome, studies with larger sample sizes are more powerful—they can detect smaller impacts on that outcome. Power is calculated with standard formulas and commonly represented as a *minimum detectable effect size*, which is the effect that will be statistically significant with a probability conventionally set to 80 percent.

For the reading test, the study obtained responses from 571 treatment group students and 366 control group students in the third year. This yielded a minimum detectable effect size of 0.13, which translates into a difference between the treatment and control groups of 5 percentile points (table B-1). For parent-reported school safety, the study obtained responses from 504 treatment group parents and 360 control group parents, which yielded a minimum detectable effect size of 0.18 that translates into a difference of 9 percentage points. For student-reported school safety, the study obtained responses from 364 students in the treatment group and 220 students in the control group—this sample included only students in grades 4 or higher. The minimum detectable effect size was 0.20, equivalent to an increase of 10 percentage points.

Table B-1. Minimum detectable effect sizes

Outcome	Treatment group sample size at followup	Control group sample size at followup	Minimum detectable effect size	Impact in units of the outcome
Reading score	571	366	0.13	5 percentile points
Student-reported school safety	364	220	0.20	10 percentage points
Parent-reported school safety	504	360	0.18	9 percentage points
Percent of parents giving school a grade of A or B	517	368	0.17	8.5 percentage points
Parent involvement with schools	474	345	0.17	2 events
Reading score				
Subgroup				
SINI	401	217	0.17	7 percentile points
Not SINI	170	149	0.22	9 percentile points
Student is below median in reading	282	165	0.19	7.5 percentile points
Student is above median in reading	289	201	0.18	7 percentile points
Elementary school students	400	278	0.16	6 percentile points
Middle/high school students	171	88	0.26	10.5 percentile points
Percent of parents giving school a grade of A or B				
Subgroup				
SINI	378	234	0.20	10 percentage points
Not SINI	139	134	0.30	15 percentage points
Student is below median in reading	266	174	0.24	12 percentage points
Student is above median in reading	251	194	0.24	12 percentage points
Elementary school students	341	240	0.20	10 percentage points
Middle/high school students	176	128	0.29	14.5 percentage points

SOURCE: OSP applications, *TerraNova Third Edition* reading and mathematics tests, parent and student surveys for OSP evaluation, and author's calculation.

Table B-1 also shows detectable effects for two outcomes and three subgroups. (Detectable effects for mathematics subgroups would be nearly the same as for reading subgroups and are not shown here.) The table shows that within subgroups, detectable effect sizes ranged from 0.16 to 0.30. For test scores, the effect sizes were equivalent to students moving 6 to 10.5 percentile points (for example, from the 50th percentile to the 56th or 44th percentile). For percentage of parents giving a school a grade of A or B, it meant the treatment group average needed to be 10 to 15 percentage points different from the control group average.

B-3. Estimating Impacts

The study's approach for estimating impacts was to model an outcome after application to the OSP (e.g., mathematics achievement) as a function of student baseline (pre-OSP) test scores and student and parent characteristics (all of which were covariates in the model), and whether the student received an offer of a scholarship. This estimate is referred to as the *intent-to-treat* impact. The offer of a scholarship created an "intent" for a student to be treated, which in this context means using the scholarship to attend a participating private school. The study used the intent-to-treat impact as a basis for estimating the impact of using the scholarship, referred to as the *treatment-on-treated* impact.

Because eligible applicants to the OSP were randomly assigned by the lottery, on average, the treatment and control groups of students should be identical at the time of the lottery, which allows the study to attribute differences in average outcomes to receiving a scholarship offer. In practice, small differences in characteristics such as academic achievement and demographic background can arise. Also, reducing variances of outcomes yields more statistical power, as noted above. For these reasons, conventional practice is to use linear regression models to estimate impacts.

The structure of regression models used here is shown in equation (1):

$$S_{it} = \alpha + \beta T_i + X_{i0}\Gamma + \delta READ_{i0} + \eta MATH_{i0} + \theta Days_{it} + \varepsilon_{it} \quad (1)$$

S_{it} is the test score for student i in year t. The time of application is 0, the baseline, and two years later is $t = 2$, which is when the outcomes were measured for this report. T_i is a (0,1) indicator indicating whether the student was in the treatment group (received a scholarship offer). It is fixed by the lottery, so it does not have a time dimension. The key coefficient in this model is β, which measured the impact of receiving a scholarship offer on the outcome of interest. X_{i0} is a set of student characteristics measured at time 0, and $READ_{i0}$ and $MATH_{i0}$ are reading and mathematics scores measured at time 0. Students were tested in their home schools, and timing of these tests varied between students, which was accounted for in the regression by including a variable $Days_{it}$

The model included the following covariates:

- Indicator for year of application (spring 2012, 2013, or 2014)

- Indicator for grade level the child was entering the next school year
- *TerraNova* test scores in reading and mathematics at the time of application
- Number of days from September 1 to date of followup test
- Indicator for whether student was enrolled in a SINI school at time of application
- Student demographic characteristics (gender, race, disability, age difference from median age for grade)
- Family characteristics (employment, college education, income, number of children, months at current address)
- Parent's rating of safety and satisfaction with child's school at time of application[6]

A variant of the model was used to estimate impacts for the safety and satisfaction outcomes. These outcomes had a value of either 0 or 1 and required different estimation techniques than for test scores, but the models included the same covariates.[7]

Main Models

To provide additional detail about the study's model, table B-2 presents impact estimates for the full models of reading and mathematics outcomes. While the report focuses on the impact of the scholarship offer on achievement outcomes, the table shows how each covariate included in the model is associated with the outcomes being measured. The coefficients suggest that test scores at the time of application were highly predictive of later achievement scores. For example, increase of one point in the reading test score at the time of application is associated with about a third of a point increase (0.32) in reading achievement three years later. Other coefficients followed intuitive patterns. For example, students with disabilities scored lower on average, and students in higher-income families scored higher on average (though all families would be considered low-income to be eligible for the program).

[6] Even parents of pre-K students completed ratings of safety and satisfaction with their child's current school at time of application. These students may have been in traditional public school preschools, private schools, or very different settings, including home daycare.
[7] Although impacts on "binary" outcomes (those that take on only two values) are often estimated using logistic models, researchers increasingly use linear probability models because in practice they yield the same results but the results are easier to interpret. The study estimated and compared both types of models and found the same direction of results and levels of statistical significance.

Table B-2. Impact estimates for full model of reading and mathematics tests three years after application

Characteristic	Reading		Mathematics	
	Impact estimate	*p*-value	Impact estimate	*p*-value
Treatment	-1.76	0.46	0.17	0.96
Year of application				
Second cohort (spring 2013)	0.29	0.92	-5.26	0.25
Third cohort (spring 2014)	-4.13	0.30	-5.35	0.37
Entering grade				
Grade 1	-15.91	<0.01	0.79	0.92
Grade 2	-30.08	<0.01	-17.99	0.05
Grade 3	-24.30	<0.01	-14.32	0.14
Grade 4	-37.25	<0.01	-22.88	0.04
Grade 5	-29.32	<0.01	-24.01	0.04
Grade 6	-34.28	<0.01	-2.56	0.83
Grade 7	-38.16	<0.01	-21.53	0.14
Grade 8	-46.16	<0.01	-11.89	0.46
Grade 9	-36.65	<0.01	-6.49	0.66
Grade 10	-48.30	<0.01	-13.88	0.39
Test score				
Reading scale score at time of application	0.32	<0.01	0.28	<0.01
Mathematics scale score at time of application	0.17	<0.01	0.30	<0.01
Student characteristics				
Student is female	6.83	0.01	-6.08	0.11
Student is African American	-1.87	0.52	-2.67	0.59
Student has disabilities or other challenges	-9.73	0.02	-13.91	0.03
Student attends a school in need of improvement	9.57	0.02	-6.11	0.24
Student age difference from median age of grade	-3.03	0.25	-1.93	0.69
Days from September 1 to followup test	-0.10	0.15	0.08	0.44
Family characteristics				
Parent went to college	-1.08	0.66	-0.49	0.90
Parent gave school grade of A or B at time of application	4.21	0.10	8.38	0.06
Parent perception of school safety at time of application	-4.02	0.14	-8.69	0.04
Parent was employed at time of application	-5.62	0.03	-4.39	0.31
Family income in thousands at time of application	0.10	0.31	0.38	0.01
Number of children in household at time of application	0.19	0.81	1.89	0.17
Months at current address at time of application (in tens)	-0.34	0.09	-0.24	0.35
R^2	0.51		0.47	

NOTE: Sample size was 937 students for reading and 934 students for mathematics.

SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered three years after application. Estimated impacts were generated from the study's regression models, as described in appendix section B-3.

Alternative Models Considered

A classical regression model assumes random errors between any two participants are uncorrelated. However, some students in the OSP sample are siblings in the same families, and it is unlikely their random errors are uncorrelated. The study's approach was to estimate impacts using "generalized estimating equations," with families specified as a group variable (on generalized estimating

equations, see Liang and Zeger [1986]). This approach was consistent with the clustering approach the first OSP study used (see Wolf et al. 2010) and was selected for the current study both to maintain comparability and because family-level clustering is a more conservative analysis strategy than other alternatives that were considered, such as clustering by school. The first impact report for the current study (Dynarski et al. 2017) compared effects that clustering had on variances and found that allowing for family clustering in estimating impacts on reading and mathematics test scores resulted in variances being larger by 3.1 percent for reading and 2.8 percent for mathematics. Allowing for school clustering resulted in variances being 1.3 percent smaller for reading and 1.7 percent larger for mathematics.

An alternate approach to estimation involves using higher-order terms (e.g., a cubic function) in the models (see Chingos and Kuehn 2017). Using a polynomial model to estimate impacts for reading and mathematics found that neither of the higher-order terms was statistically significant, and impacts were similar to the primary model (table B-3).

Table B-3. Comparison of model estimates from primary regression, polynomial, and zero-value replacement of the impacts of offering a scholarship on reading and mathematics achievement three years after application

	Primary model		Polynomial model		Zero-value replacement for missing indicators model	
Outcome	Impact estimate	*p*-value	Impact estimate	*p*-value	Impact estimate	*p*-value
Reading achievement	-1.76	0.46	-1.48	0.53	-0.66	0.96
Mathematics achievement	0.17	0.96	0.06	0.99	-2.32	0.48

NOTE: Sample size for primary and polynomial models was 937 students for reading and 934 students for mathematics. Sample size for the zero-replacement model was 1,182 for reading and 1,179 for mathematics.

Table B-3 also presents results for an alternate approach to cases with missing covariates. Students and parents were dropped from the estimation models if any of their covariates were missing, which is termed a "complete case" analysis. An alternate approach is to leave these cases in the sample as a zero value and add a flag to the model to indicate that the zero value was replacing missing data for some covariate, which we term "zero-value replacement." A student missing a baseline test score, for example, would have a value of zero inserted for their test score and the flag for missing test scores would be set to a value of 1. (A student that had a baseline test score would have that score in the model and the flag for missing test scores would be set to 0.) Comparing this approach to the primary model, the impact for reading was slightly negative and insignificant for both models. The impact for mathematics changed from a small positive impact to a small negative impact but again was insignificant.

Estimating Subgroup Impacts

For subgroup analyses, equation (1) above was modified to allow for an interaction between the indicator for students in the treatment group and an indicator for membership of a given subgroup. The model included an interaction between the subgroup indicator and treatment, and the subgroup indicator

was included as an additional explanatory variable. This ensured that the coefficient on the interaction was not picking up a direct relationship between the outcome variable and the subgroup indicator. The equation below assumes that the entire sample was divided into two groups, with G_i an indicator for whether student *i* belongs to the particular group.

$$(2) \qquad S_{it} = \alpha + \beta T_i + \pi G_i + \rho G_i T_i + X_{i0}\Gamma + \delta READ_{i0} + \eta MATH_{i0} + \theta Days_{it} + \varepsilon_{it}$$

In this equation, β measures the impact for the omitted subgroup (those not in group G), ρ captures the *difference* between the impact on the omitted group and group G, and the sum $\beta + \rho$ captures the estimate of the total impact of treatment for group G. For outcomes other than test scores, the same modification was made to equation (2) to allow for the relationship between the given outcome and both group G and the interaction between G and treatment status.

Estimating Impacts of Using a Scholarship

The Scholarships for Opportunity and Results (SOAR) Act specifies that the evaluation measure both the impact of being offered a scholarship and the impact of *using* a scholarship. This latter impact, sometimes called the impact of "treatment on the treated" (TOT), can be estimated in a straightforward way by dividing the impact of being offered a scholarship by the fraction of the treatment group that uses the scholarship (Bloom 1984). For example, if an impact of the offer were estimated to be 10 points, and half of the treatment group used their scholarship, the impact of using a scholarship would be estimated to be 20 points (10 divided by 50 percent). This adjustment relies on the assumption that students are not affected by the offer unless they use their scholarship. This assumption would be violated if the offer changed student or family behavior in some way that affected outcomes even if the scholarship were not used. Other approaches to estimating the impacts of using a scholarship have been developed, but in practice tend to yield similar estimates (Angrist, Imbens, and Rubin 1996). A comparison of TOT estimates using the Bloom adjustment with estimates from an instrumental variables (IV) approach was conducted for this study's first impact report. The two methods produced very similar estimates (table B-4).

Table B-4. Comparison of Bloom adjustment and instrumental variables estimates of the impacts of using a scholarship (TOT estimates) on reading and mathematics achievement in Year 1

	Bloom adjustment		Instrumental variables		Difference of
Outcome	TOT estimate	*p*-value	TOT estimate	*p*-value	estimates
Reading achievement	-5.42	0.12	-5.48	0.13	0.06
Mathematics achievement	-8.92	0.03	-8.96	0.04	0.04

In the third year, there are six semesters in which students could have used their scholarship. The study defined "use" to be any use in the six semesters.

B-4. Method for Calculating Percentile Changes

Scale scores from standardized tests are useful in regression models because of their statistical properties, but they can be difficult to interpret. Percentile changes are easier to interpret, but because of the study's K–12 grade range, converting scale scores to percentile changes required additional considerations discussed here.[8] The considerations center on the fact that students in different grade levels were in different places relative to the national distribution. Students in lower grade levels were higher in the distribution than students in higher grade levels.

Impacts were depicted as the difference in average percentiles for the treatment group and the control group. The overall percentile difference was found by computing percentile differences at each grade level, and then weighting those differences by the proportion of the student sample at each grade level. The approach to compute percentile changes has three steps:

1. At each grade level, the average scale score for the control group was compared to the national *TerraNova* score distribution for that grade level. The average was converted to a percentile of the national distribution using a quantile function, in this case the inverse normal cumulative distribution function. Grades scoring above the national average had percentiles greater than 50, and grades scoring below the national average had percentiles less than 50.

2. At each grade level, the average scale score for the treatment group was computed as the average scale score for the control group plus the estimated treatment impact, which was assumed to be the same for each grade level. For example, the average reading score for second grade students in the control group was 594, which put these students at the 45th percentile relative to the national sample. The average score for second grade students in the treatment group was 594 of the control group minus the impact of 1.76 points, which yielded a score of 592 and put these students at the 44th percentile, relative to the national sample.[9]

3. Steps (1) and (2) yielded 11 differences between percentiles of the treatment and control groups (table B-5). These differences were averaged using the proportion of the sample at each grade level as weights.

This procedure yielded a negative percentile change if the impact on scores was negative, and vice versa. However, the same magnitude of the score impact had different effects on percentile changes depending on the grade level. The same procedure was used for student subgroup results presented in this report.

[8] The study also considered using *z*-scores, which used scale scores at each grade level and then adjusted them to have a mean of zero and a standard deviation of one. However, the *TerraNova* does not include national-norm information for entering kindergartners, a large component of the study's sample. And *z*-scores do not have a direct interpretation, and ultimately would need to be converted to percentile differences to be interpretable.

[9] The model estimated an overall impact, which applied to all students in the sample, and that overall impact was used to calculate percentile changes. In theory, grade-level impacts could be used to calculate percentile changes, but these would be highly variable because of the small samples in each grade.

Table B-5. Computing percentile changes in reading scores, by grade level

Grade	OSP control group mean scale score	*TerraNova* national mean scale score	*TerraNova* national standard deviation	OSP control group mean as percentile	OSP treatment group mean as percentile	Change of percentile
2	594	599	42	45	44	-2
3	614	622	39	42	41	-2
4	622	637	39	35	34	-2
5	645	652	39	43	42	-2
6	645	658	41	37	36	-2
7	655	664	41	41	40	-2
8	664	674	40	40	38	-2
9	654	679	41	27	26	-1
10	661	688	43	26	25	-1
11	684	700	44	35	34	-1
12	660	708	44	14	13	-1

SOURCE: National mean and standard deviation from *TerraNova Third Edition Technical Report* (CTB/McGraw-Hill 2010). Estimated OSP means were generated from the study's regression models, as described in appendix section B-3.

B-5. Outcome Measures and Data Collection Procedures

To estimate impacts, the study collected data on outcomes and characteristics of students, parents, and schools from a variety of sources (table B-6). The program required parents (or guardians) to complete an application form to apply for a scholarship,[10] and the application process included baseline (pre-program) testing of students in reading and mathematics by the evaluation team. As a result, the study had nearly complete data about students and families at the time of application. Parents were surveyed and students were surveyed and tested each year after the initial application.

Table B-6. Data sources used to estimate impacts

Outcome	Source
Student achievement in reading and mathematics	*TerraNova Third Edition*
Parent satisfaction with school Parent perceptions of school safety Parent involvement with education at school Parent involvement with education in the home	Parent survey
Student satisfaction with school Student perceptions of school safety	Student survey, grades 4–12
Student chronic absenteeism	Administrative records and private school student outcome records form

Student achievement in reading and mathematics. For its academic achievement outcome, the study used reading and mathematics tests from the CTB/McGraw-Hill *TerraNova Third Edition* (CTB/McGraw-Hill 2008).[11] These nationally normed standardized tests are vertically aligned and

[10] Parents were asked to complete all application questions, and parents of pre-K students responding to survey items about satisfaction with their child's school and perceptions of school safety may have been providing ratings for a range of settings including public preschool or home daycare.

[11] DC administers its own standardized assessment in grades 3 through 8 and, during the early years of the evaluation, was administering an assessment in grade 10. However, aspects of the study precluded using these test scores for this study: the OSP statute required the evaluation to use a nationally normed assessment (the DC one is not), private schools do not need to use the DC assessment, and the study had students in the entire K–12 grade range, which included grades that do not administer the DC assessment.

available for grades K–12. The study selected the *TerraNova, Third Edition* assessment because the abbreviated battery, which is available for grades 2–12, offered shorter test administration times for most students (about 90 minutes).

Students were tested at the time of application, which provided a baseline test score that was used as an adjustment variable in estimating impacts.[12] Followup testing was conducted at the schools where students were enrolled in the spring of each year following application. For this report, which examines impacts three years after being offered or using a scholarship, testing took place during spring 2015 for the first cohort, in 2016 for the second cohort, and in 2017 for the third cohort (table B-7). Annual testing was conducted with students at the school they were attending in spring of the second year after applying to the program. The spring data collection window was designed to occur as close to two years after baseline testing as possible. The study worked with school staff members to schedule times and locations for the assessments that minimized disruption for students. Students in grades K–2 were tested in groups of 5 or fewer, while students in grades 3–12 were tested in groups of 10 or fewer. Limiting the time to administer the test was critical to ensuring school cooperation with the study's data collection effort.

The study used trained staff to administer the *TerraNova* student assessments in reading and mathematics, using the full battery for grades K–1 and abbreviated batteries available for grades 2–12. Test administrators attended annual trainings before the start of each data collection period. A representative from the test publisher (CTB/McGraw-Hill) trained study staff on test administration procedures and standardized testing protocols. The staff followed the test publisher's scripts and instructions during testing to ensure that testing conditions were similar across all schools in the study to minimize potential bias.

Table B-7. Study cohorts and years tested

Cohort	Spring 2012	Spring 2013	Spring 2014	Spring 2015	Spring 2016	Spring 2017
1	Application and lottery	Data Collection 1	Data Collection 2	Data Collection 3		
2		Application and lottery	Data Collection 1	Data Collection 2	Data Collection 3	
3			Application and lottery	Data Collection 1	Data Collection 2	Data Collection 3

The *TerraNova, Third Edition* uses multiple-choice questions to measure subject area content and process skills. For grades K–2, the test focuses on the basic concepts of number, operations, measurement, geometry, patterns, and data representation. For grades 3–5, the test focuses on estimation, probability, simple functions, and inferences from data. For grades 6–12, the test covers more advanced applications of the basic concepts and data presentations, statistics, graphs, and problem solving

[12] Random assignment yields student groups who are equivalent in theory, but measuring achievement at the time of application added considerable statistical power to the estimation and adjusted for differences between treatment and control groups that arose due to chance variation.

situations. The reading test in grades K–2 includes oral (listening) comprehension, word analysis skills, phonics, and phonemic awareness. In the later primary and secondary grades, the focus is on reading comprehension using informational, narrative, expository text selections.

The *TerraNova's* vertical scaling allowed the OSP evaluation to analyze scores from students in different grade levels (i.e., K–12) in the same model. The test publisher administered test forms with common items to respondents in each pair of adjacent grade levels. The publisher used a procedure established by Stocking and Lord (1983) to equate scores from one grade to those of the adjacent grade, creating a vertical scale across grades.

Absenteeism. The number of days students were absent came from Office of State Superintendent for Education records and the student records forms collected from private schools for participating students. The number of days in the school year was compiled for each school using public records. This information was then used to convert the number of days a student was absent to a percent of the school year absent.

Student surveys. Students in grades 4–12 completed a brief survey immediately after completing the assessment. The student survey provided outcome measures for student satisfaction and perceptions of safety. Other topics included attitude toward school, school environment, friends and classmates, and involvement in activities.

Parent surveys. Parent surveys provided self-reported outcome measures for parent satisfaction, perceptions of school safety, and parental involvement in education at school and in the home. A parent or guardian was asked to complete a brief survey for each child in their family who applied for an OSP scholarship. Each year, parents were contacted by mail and email to request that they complete the online survey. Parents were provided links and access codes for the web-based survey and paper copies were provided in followup mailings. The study also conducted followup calls to nonrespondents and offered the option to complete the survey with an interviewer by phone. Parents who completed the survey received a modest payment.

Tables B-8 through B-10 describe response rates for student tests, parent surveys, and student surveys in the third year of followup. These respondents constitute the analysis samples for this report.

Table B-8. Student test response rates for third-year followup

Group	Original sample*	Reading respondents	Reading response rate (%)	Mathematics respondents	Mathematics response rate (%)
All students	1,725	1,182	68.5	1,179	68.3
Treatment group	968	712	73.6	710	73.3
Control group	757	470	62.1	469	62.0

* Of the original 1,771 students, 46 were entering grades 11 or 12 at the time of application and were no longer part of the study's data collection in the third year.

SOURCE: *TerraNova Third Edition* reading and mathematics tests.

Table B-9. Parent survey response rates for third-year followup

Group	Original sample	Respondents	Parent response rate (%)	Parent effective respondents	Effective response rate (%)*
All students	1,725	1,095	63.5	1,205	69.9
Treatment group	968	642	66.3	677	70.0
Control group	757	453	59.8	528	69.7

* Response rates were increased through the use of subsampling, which is described in more detail in appendix section B-7.
SOURCE: Parent surveys for OSP evaluation, 2014–2016.

Table B-10. Student survey response rates for third-year followup

Group	Original sample*	Respondents	Student response rate (%)
All students	1,091	687	63.0
Treatment group	625	424	67.8
Control group	466	263	56.4

* Students in grades 4 and above in the third year.
SOURCE: Student surveys for OSP evaluation, 2015–2017.

Other data sources. Application data and payment files documenting student's use of the scholarship was provided by the OSP program operator. Information about tuition rates for OSP participating private schools was obtained from the OSP school directories published by the program operator. Data on the public school characteristics that students in the study sample attended were obtained from the National Center for Education Statistics (NCES) Common Core of Data. Data on the characteristics of private schools was obtained from the NCES Private School Survey.

Table B-11 presents rates of missing data for the study's key outcomes and covariates in the third year of data collection. For example, 26 percent of reading scores were missing for the treatment group in the third followup year, and 38 percent of reading scores were missing for the control group in that year. (Appendix D presents analyses of the extent to which these differential rates of missingness may have affected the findings.) Students who were missing data on outcomes or covariates at the time of application were dropped from the analysis (as described in appendix section B-3, an alternative approach for handling missing data found similar results).

Table B-11. Sample size, valid sample, and percentage missing data at third-year followup

	Treatment			Control		
	Sample size	Non-missing sample size	Percent missing	Sample size	Non-missing sample size	Percent missing
Outcomes						
Reading score	968	712	26	757	470	38
Mathematics score	968	710	27	757	469	38
Student reported satisfaction	625	412	34	466	253	46
Student reported safety	625	406	35	466	255	45
Parent overall satisfaction with child's school	968	637	34	757	446	41
Parent reported safety of school	968	624	36	757	439	42
Frequency of parent educational activities	968	625	35	757	436	42
Frequency of parent communications with school	968	589	39	757	416	45
Covariates						
Gender	968	968	0	757	757	0
Race	968	968	0	757	757	0
Reading score at time of application	968	941	3	757	729	4
Mathematics score at time of application	968	924	5	757	708	6
Attending a school in need of improvement	968	968	0	757	757	0
Whether student has a learning disability	968	968	0	757	757	0
Whether student has an individual education program (IEP)	968	968	0	757	757	0
Parent's education	968	964	<1	757	750	1
Parent's employment status	968	964	<1	757	750	1
Household income	968	968	0	757	757	0
Number of children in household	968	957	1	757	751	1
Number of months at current address	968	955	1	757	749	1
Parent satisfaction with school	968	843	13	757	673	11
Parent satisfaction with school safety	968	864	11	757	684	10
Days from September 1 to followup test	968	714	26	757	471	38

NOTE: Of the original 1,771 eligible applicants, 46 were entering 11th or 12th grade at the time of application and were no longer part of the study's data collection in the third year. This table shows data available to measure key outcomes and student/family characteristics (covariates in the study's models) for the 968 treatment group and 757 control group students in the third year.

SOURCE: OSP applications, *TerraNova Third Edition* reading and mathematics tests, parent and student surveys for OSP evaluation.

B-6. Baseline Characteristics for Third-Year Impact Sample

Tables B-12 through B-14 present baseline characteristics for the samples of students and parents who completed tests and surveys in the third year of data collection. Table B-12 shows statistically significant differences between the treatment group and control group for five characteristics. As described in appendix section B-3, the study's statistical models adjust for these differences when estimating impacts. Fewer significant differences were observed for the sample of students whose parents completed the parent survey (table B-13), and for the sample of students who completed the student survey (table B-14), administered to students in grades 4–12.

Table B-12. Characteristics of treatment and control groups at time of application, for students who completed reading tests three years after application

	Treatment			Control			
Characteristic	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference of means
Year of application							
First cohort (spring 2012)	571	28.4%	45.1%	366	27.4%	44.6%	1.0
Second cohort (spring 2013)	571	42.8	49.5	366	43.5	49.6	-0.7
Third cohort (spring 2014)	571	28.8	45.3	366	29.1	45.4	-0.3
Entering grade							
Kindergarten	571	18.2%	38.6%	366	21.7%	41.2%	-3.5
Grade 1	571	13.5	34.2	366	12.7	33.3	0.8
Grade 2	571	10.7	30.9	366	10.2	30.2	0.5
Grade 3	571	11.0	31.3	366	9.1	28.7	1.9
Grade 4	571	8.0	27.1	366	9.5	29.3	-1.5
Grade 5	571	7.1	25.8	366	6.3	24.2	0.9
Grade 6	571	12.1	32.6	366	8.1	27.3	4.0*
Grade 7	571	6.3	24.4	366	3.1	17.4	3.2*
Grade 8	571	3.7	18.9	366	6.3	24.4	-2.6
Grade 9	571	7.2	25.8	366	9.2	29.0	-2.1
Grade 10	571	2.1	14.3	366	3.8	19.1	-1.7
Test score							
Reading scale score at time of application	571	570.0	82.2	366	567.3	89.9	2.7
Grades K–2	249	493.6	53.0	182	487.1	57.5	6.5
Grades 3–5	151	599.8	45.5	96	602.9	50.4	-3.1
Grades 6–8	123	637.2	43.5	54	641.3	28.1	-4.1
Grades 9–10	48	664.1	41.2	34	675.0	29.0	-10.9
Mathematics scale score at time of application	571	541.5	108.0	366	542.6	112.8	-1.1
Grades K–2	249	444.4	68.3	182	444.2	63.0	0.2
Grades 3–5	151	569.0	62.6	96	577.6	64.2	-8.6
Grades 6–8	123	629.5	56.0	54	632.1	61.4	-2.6
Grades 9–10	48	685.3	45.1	34	682.4	51.9	2.9
Student characteristics							
Student is female	571	49.8%	50.0%	366	49.8%	50.0%	0.1
Student is African American	571	84.9%	35.8%	366	86.9%	33.7%	-2.0
Student has disabilities or other challenges	571	14.7%	35.4%	366	9.3%	29.0%	5.4*
Student attends a school in need of improvement	571	70.0%	45.8%	366	67.1%	47.0%	2.9
Student age difference from median age of grade	571	<0.1	0.4	366	<-0.1	0.5	<0.1

See notes at end of table.

Table B-12. Characteristics of treatment and control groups at time of application, for students who completed reading tests three years after application—Continued

	Treatment			Control			
Characteristic	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	Difference of means
Family characteristics							
Parent went to college	571	59.6%	49.1%	366	58.4%	49.3%	1.2
Parent gave school grade of A or B at time of application	571	59.2%	49.1%	366	58.1%	49.3%	1.2
Parent perception of school safety at time of application	571	75.4%	43.1%	366	68.3%	46.5%	7.1*
Parent was employed at time of application	571	47.4%	49.9%	366	44.6%	49.7%	2.8
Family income in thousands at time of application	571	11.8	12.7	366	13.2	13.0	-1.4
Number of children in household at time of application	571	2.5	1.3	366	2.7	1.4	-0.2*
Months at current address at time of application (in tens)	571	6.9	8.5	366	6.1	7.5	0.8

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: This table shows baseline characteristics for the 571 students in the treatment group, and 366 students in the control group who completed the reading achievement test in the third year of followup. Three students completed the reading but not the mathematics achievement test, so the analysis sample for mathematics outcomes was very similar. For binary variables (e.g., grade level or female), the mean is the proportion of positive responses, and the standard deviation measures how spread out the distribution is from that proportion.

SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered at time of application.

Table B-13. Characteristics of treatment and control groups at time of application, for parents who completed surveys three years after application

Characteristic	Treatment			Control			Difference of means
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	
Year of application							
First cohort (spring 2012)	517	29.5%	45.6%	368	28.6%	45.2%	0.9
Second cohort (spring 2013)	517	41.8	49.3	368	41.7	49.3	0.1
Third cohort (spring 2014)	517	28.8	45.3	368	29.8	45.7	-1.0
Entering grade							
Kindergarten	571	16.2%	36.9%	368	19.4%	39.6%	-3.2
Grade 1	517	12.3	32.9	368	10.7	30.9	1.6
Grade 2	517	9.4	29.2	368	8.0	27.2	1.4
Grade 3	517	12.5	33.1	368	9.5	29.3	3.0
Grade 4	517	8.9	28.5	368	8.5	27.9	0.4
Grade 5	517	7.5	26.3	368	7.4	26.2	<0.1
Grade 6	517	10.9	31.2	368	7.8	26.9	3.1
Grade 7	517	5.5	22.8	368	6.2	24.1	-0.7
Grade 8	517	5.1	22.0	368	7.1	25.7	-2.0
Grade 9	517	7.4	26.1	368	9.1	28.8	-1.8
Grade 10	517	4.2	20.2	368	6.1	23.9	-1.8
Test score							
Reading scale score at time of application	517	572.3	85.1	367	574.3	88.9	-2.0
Mathematics scale score at time of application	517	546.7	105.8	368	551.9	107.6	-5.2
Student characteristics							
Student is female	517	51.7%	50.0%	368	52.6%	49.9%	-0.9
Student is African American	517	86.5%	34.2%	368	85.9%	34.8%	0.6
Student has disabilities or other challenges	517	15.9%	36.6%	368	12.0%	32.5%	3.9
Student attends a school in need of improvement	517	71.0%	45.4%	368	67.1%	47.0%	4.0
Student age difference from median age of grade	517	<0.1	0.5	368	0.7	0.5	-0.7
Family characteristics							
Parent went to college	517	60.1%	49.0%	368	62.9%	48.3%	-2.8
Parent gave school grade of A or B at time of application	517	58.3%	49.3%	368	57.7%	49.4%	0.6
Parent perception of school safety at time of application	517	74.5%	43.6%	368	69.2%	46.1%	5.3
Parent was employed at time of application	517	46.8%	50.0%	368	44.2%	49.7%	2.6
Family income in thousands at time of application	517	12.1	12.5	368	12.1	12.5	<0.1
Number of children in household at time of application	517	2.5	1.4	368	2.8	1.4	-0.3*
Months at current address at time of application (in tens)	517	7.2	8.5	368	6.1	7.4	1.1*

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: This table shows baseline characteristics for the 517 students in the treatment group and the 368 students in the control group whose parents completed the parent survey in the third year of followup. For binary variables (e.g., grade level or female), the mean is the proportion of positive responses, and the standard deviation measures how spread out the distribution is from that proportion.

SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered at time of application.

Table B-14. Characteristics of treatment and control groups at time of application, for students who completed surveys three years after application

Characteristic	Treatment			Control			Difference of means
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation	
Year of application							
First cohort (spring 2012)	368	31.4%	46.4%	219	27.5%	44.6%	3.9
Second cohort (spring 2013)	368	42.5	49.4	219	42.8	49.5	-0.3
Third cohort (spring 2014)	368	26.2	44.0	219	29.8	45.7	-3.6
Entering grade							
Grade 2	368	15.6	36.3	219	15.1	35.8	0.5
Grade 3	368	16.2	36.8	219	12.9	33.6	3.2
Grade 4	368	11.8	32.3	219	16.1	36.8	-4.3
Grade 5	368	10.5	30.7	219	11.5	31.9	-0.9
Grade 6	368	17.5	38.0	219	12.9	33.6	4.6
Grade 7	368	10.6	30.8	219	5.9	23.6	4.7*
Grade 8	368	5.0	21.9	219	9.2	28.9	-4.2*
Grade 9	368	10.1	30.2	219	11.7	32.2	-1.6
Grade 10	368	2.6	16.0	219	4.5	20.8	-1.9
Test score							
Reading scale score at time of application	368	611.3	56.4	219	617.7	54.0	-6.4
Mathematics scale score at time of application	368	593.9	76.9	219	597.2	82.3	-3.3
Student characteristics							
Student is female	368	50.3%	50.0%	219	54.7%	49.8%	-4.4
Student is African American	368	85.5%	35.2%	219	86.8%	33.9%	-1.3
Student has disabilities or other challenges	368	16.0%	36.7%	219	11.5%	31.9%	4.5
Student attends a school in need of improvement	331	87.3%	33.3%	219	85.7%	35.0%	1.6
Student age difference from median age of grade	368	<0.1	0.5	219	<-0.1	0.5	0.1
Family characteristics							
Parent went to college	368	56.7%	49.5%	219	58.9%	49.2%	-2.2%
Parent gave school grade of A or B at time of application	368	56.3%	49.6%	219	54.9%	49.8%	1.5%
Parent perception of school safety at time of application	368	74.6%	43.5%	219	71.3%	45.3%	3.4%
Parent was employed at time of application	331	47.5%	49.9%	219	44.0%	49.6%	3.5%
Family income in thousands at time of application	368	11.9	12.6	219	12.8	13.4	-1.0
Number of children in household at time of application	368	2.5	1.3	219	2.7	1.4	-0.2
Months at current address at time of application (in tens)	368	6.9	8.5	219	6.0	7.3	0.9

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: This table shows baseline characteristics for the 331 students in the treatment group and 186 students in the control group who completed the student survey in the second year of followup. For binary variables (e.g., grade level or female), the mean is the proportion of positive responses, and the standard deviation measures how spread out the distribution is from that proportion.

SOURCE: OSP applications and *TerraNova Third Edition* reading and mathematics tests administered at time of application.

B-7. Sampling and Nonresponse Weights

Weights were used in estimating impacts to offset the different probabilities that some applicants had in the lottery and to adjust for nonresponse. Weights had two parts: (1) a "base weight," which is the inverse of the probability of being selected to treatment (or control), and (2) an adjustment for differential nonresponse.

Constructing Base Weights

The base weight is the inverse of the probability of being assigned to either the treatment or control group. For each randomization stratum *s* defined by cohort, SINI status, and sibling status, *p* is the probability of assignment to the treatment group (receiving an offer of a scholarship) and *1-p* the probability of being assigned to the control group.

Adjustments for Nonresponse

The initial base weights were adjusted for nonresponse, where a "respondent" was of five types: (i) a student who had completed a *TerraNova* reading or mathematics test, (ii) a parent who had completed the questionnaire, (iii) a student who had completed the questionnaire, (iv) a student who had attendance data, and (v) a student whose principal had completed a questionnaire. The use of these weights helped control bias by compensating for different response rates across groups of students or parents. Essentially, nonresponse weights put more weight on students or parents that "look like" nonresponding students or parents.

The approach taken to constructing nonresponse-adjusted weights is based on a "pseudo-randomization" framework in which respondents are treated as a stratified random sample from the full sample . This underlying, unknown pseudo-sampling rate is called a response propensity. See for example Lohr (1999), Section 8.4. An early reference for this is Oh and Scheuren (1983). This approach will yield unbiased estimates if the data are "missing at random (MAR)", meaning that the response propensity is independent of the outcome variable conditional on the set of baseline auxiliary variables known for all members of the sample and used to construct the weights. See for example Little and Rubin (1987).

To construct the weights, we estimated a model of nonresponse. The baseline variables considered for inclusion in the nonresponse model were family income, parent or guardian's job status, parent or guardian's education, length of time at current address, disability status of the child, race, grade, gender, and baseline test score data (both reading and mathematics). To select the subset of these variables for inclusion in the nonresponse model, we applied stepwise logistic regression with a p-value threshold set to 0.20 (20 percent). These stepwise procedures were performed separately within each sampling stratum. The study then created nonresponse adjustment cells, and within cells used the Chi-squared Automatic Interaction Detector (CHAID), approach. The CHAID program was used to identify cells with differing response rates within strata using the set of characteristics from the PROC LOGISTIC models. The nonresponse adjustment for each respondent in a cell was the reciprocal of the base-weighted

response rate within the cell. A good reference for these methods for estimating propensity to response based on baseline variables is Valliant, Dever, and Kreuter (2013), Section 13.5.

As a last step, the nonresponse-adjusted base weights were trimmed. Trimming prevents extremely large weights from inflating variances. Weights larger than 4.5 times the median weight were set to equal 4.5 times the median weight. Medians were computed separately within the treatment and control groups. See for example Valliant et al. (2013) Section 14.4.2 for a general explication of the role of trimming in weighting. An early reference is Potter (1990).

Adjusting for Nonresponse Subsampling (parent survey weights)

The study used subsampling to increase the weighted parent response rates. By subsampling 50 percent of the initial control household nonrespondents[13] then conducting intensive followup efforts with these households, the subsample allowed for a concentration of resources to improve the response outcome. See for example Cochran (1977), Section 13.6. A subsample of nonrespondents was drawn, and intensive efforts were made to get them to respond. Each initial subsampled nonrespondent who was converted to a respondent counted as one more respondent for purposes of the actual response rate, but counted as $1/(sampling\ rate_i)$ respondent for purposes of the effective response rate. The random sampling permitted respondents to "stand in" for members of the nonrespondent group who were not selected for the subsample but presumably would have converted to respondent status if they had been selected. In other words, the proportion of subsampled nonrespondents who converted represented themselves as well as the same proportion of nonsampled nonrespondents.

These "converted" cases were weighted by a factor of two (i.e., inverse of the subsampling rate or 0.5), to account for the complementary set of initial nonrespondents who were not randomly selected for targeted conversion efforts but who would have responded if they had been. The weights ensured that each converted member of the subsample represented him or herself as well as another study participant: a nonrespondent like him or her who would have converted had the person been included in the subsample.

The final student-level weights for the parent survey analysis were equal to:

$$W_i = (1/p_i) * (NR_j) * (TR_i) * (X_i)$$

where p_i is the probability of selection to treatment or control for student *i;* NR_j is the nonresponse adjustment (the reciprocal of the response rate) for the classification cell to which student *i* belongs; TR_i is the trimming adjustment (usually equal to 1, but in some cases equal to 4.5 times median cutoff divided by the untrimmed weight); and X_i is the factor for sampled nonrespondents, with X_i equal to 2.0 for this set and equal to 1 otherwise.

Tables B-15 through B-18 contain the full set of weights by study cohort and strata (priority).

[13]These were households with at least one control child without a completed survey.

Table B-15. Student reading test respondents and weights, by cohort and lottery priority

	Original sample		Respondents		Sum of base weight		Sum of final weight	
Priority/cohort	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
No priority								
Spring 2012	46	47	37	30	38.2	29.1	32.5	31.2
Spring 2013	85	102	72	69	78.6	63.6	63.6	64.5
Spring 2014	83	95	64	69	68.2	65.0	60.6	61.3
Siblings								
Spring 2012	45	22	38	10	28.3	15.2	23.0	22.9
Spring 2013	61	36	51	27	40.3	36.8	33.0	33.6
Spring 2014	43	24	39	18	30.1	25.5	22.8	23.3
SINI/Never used previous award								
Spring 2012	218	141	149	72	123.9	90.2	124.3	121.1
Spring 2013	234	180	150	108	131.6	125.5	140.7	143.3
Spring 2014	153	110	112	67	96.3	80.0	90.2	90.1
Total	*968*	*757*	*712*	*470*	*635.6*	*531.0*	*590.7*	*591.3*

SOURCE: OSP applications, *TerraNova Third Edition* reading tests.

Table B-16. Student mathematics test respondents and weights, by cohort and lottery priority

	Original sample		Respondents		Sum of base weight		Sum of final weight	
Priority/cohort	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
No priority								
Spring 2012	46	47	37	29	38.2	28.1	32.5	31.1
Spring 2013	85	102	71	69	77.5	63.6	63.4	64.3
Spring 2014	83	95	63	69	67.1	65.0	60.4	61.2
Siblings								
Spring 2012	45	22	38	10	28.3	15.2	22.9	22.9
Spring 2013	61	36	51	27	40.3	36.8	33.0	33.5
Spring 2014	43	24	39	18	30.1	25.5	22.7	23.2
SINI/Never used previous award								
Spring 2012	218	141	149	72	123.9	90.2	124.0	120.8
Spring 2013	234	180	150	108	131.6	125.5	140.4	143.0
Spring 2014	153	110	112	67	96.3	80.0	89.9	89.8
Total	*968*	*757*	*710*	*469*	*633.5*	*530.0*	*589.2*	*589.8*

SOURCE: OSP applications, *TerraNova Third Edition* mathematics tests.

Table B-17. Parent survey respondents and weights, by cohort and lottery priority

	Original sample		Respondents		Sum of base weight		Sum of final weight	
Priority/cohort	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
No priority								
Spring 2012	46	47	39	34	40.3	33.0	29.9	28.9
Spring 2013	85	102	55	55	60.1	50.7	58.9	59.7
Spring 2014	83	95	53	55	56.5	51.8	56.0	56.8
Siblings								
Spring 2012	45	22	36	13	26.8	19.8	21.5	21.3
Spring 2013	61	36	38	24	30.0	32.7	30.6	31.1
Spring 2014	43	24	30	18	23.2	25.5	21.1	19.6
SINI/Never used previous award								
Spring 2012	218	141	167	86	138.9	107.8	117.3	112.2
Spring 2013	234	180	134	109	117.6	126.7	130.4	132.8
Spring 2014	153	110	90	59	77.4	70.5	83.5	83.4
Total	*968*	*757*	*642*	*453*	*570.7*	*518.4*	*549.1*	*545.9*

SOURCE: OSP applications and parent surveys for OSP evaluation, 2015–2017.

Table B-18. Student survey respondents and weights, by cohort and lottery priority

	Original sample		Respondents		Sum of base weight		Sum of final weight	
Priority/cohort	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
No priority								
Spring 2012	16	15	10	6	10.3	5.8	10.4	9.2
Spring 2013	27	32	22	19	24.0	17.5	18.6	18.6
Spring 2014	23	29	13	18	13.9	17.0	15.5	17.2
Siblings								
Spring 2012	*	*	*	*	11.2	1.5	9.4	2.1
Spring 2013	*	*	*	*	12.6	10.9	10.0	13.7
Spring 2014	*	*	*	*	8.5	5.7	5.8	5.5
SINI/Never used previous award								
Spring 2012	171	110	115	55	95.7	68.9	89.7	87.0
Spring 2013	207	158	130	94	114.1	109.2	114.6	115.8
Spring 2014	129	98	92	58	79.1	69.3	70.0	73.9
Total	*625*	*466*	*424*	*263*	*369.4*	*305.8*	*344.0*	*343.0*

*For one or more cells, the sample size was suppressed to avoid a disclosure risk.

SOURCE: OSP applications and student surveys for OSP evaluation, 2015–2017.

Longitudinal Weights

Weights also were constructed for students who had test scores in all three years of the study. The same procedures were followed for the longitudinal weights as for the single-year weights, with some minor adjustments. Base weights for the longitudinal weights were exactly the base weights already constructed. The response-status indicator for the longitudinal weight was whether a student responded in both years, which meant the number of responders was slightly lower for the longitudinal weights than for the number of responders in each year separately. Once longitudinal status was determined, the stepwise

logistic model was run as before (for mathematics and reading separately) and the CHAID was run as before (also for mathematics and reading separately). For the previous weights, if a nonresponse adjustment factor was larger than 3.0 it was flagged for investigation, with the possibility of collapsing the nonresponse cells before proceeding. For the longitudinal weights, the flag for investigation was set at 3.5 to acknowledge the smaller sample sizes in the various cells. The trimming factor was left as 4.5.

Appendix C. Impact Findings by Outcome and Student Subgroups

This appendix provides impact estimates from the study's regression models for program outcomes in the third year, by eight student subgroups (tables C-1–C-9). Outcomes include achievement in reading and mathematics, chronic absenteeism, satisfaction, perceptions of school safety, and parent involvement.

Table C-1. Impact estimates of the offer and use of a scholarship on reading test scores after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean scale score	Control group mean scale score	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	631.68	633.44	-1.76	-0.04	-2.24	-0.04	0.46
Subgroups							
SINI	644.89	647.08	-2.19	-0.05	-2.83	-0.06	0.44
Not SINI	606.78	607.61	-0.83	-0.02	-1.03	-0.02	0.85
Difference			-1.37				0.79
Elementary school students	618.24	621.08	-2.84	-0.06	-3.48	-0.08	0.30
Middle/high school students	666.60	665.79	0.81	0.02	1.16	0.03	0.86
Difference			-3.66				0.50
Reading performance below median	611.79	616.40	-4.61	-0.09	-5.99	-0.12	0.23
Reading performance above median	650.74	649.16	1.58	0.03	1.99	0.04	0.59
Difference			-6.19				0.21
Mathematics performance below median	616.83	619.01	-2.18	-0.04	-2.78	-0.06	0.55
Mathematics performance above median	646.19	647.63	-1.44	-0.03	-1.84	-0.04	0.63
Difference			-0.75				0.87

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. *TerraNova Third Edition* reading and mathematics tests administered three years after application.

Table C-2. Impact estimates of the offer and use of a scholarship on mathematics test scores after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean scale score	Control group mean scale score	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	617.65	617.48	0.17	<0.01	0.21	<0.01	0.96
Subgroups							
SINI	635.00	632.09	2.91	0.04	3.75	0.05	0.55
Not SINI	583.53	589.21	-5.68	-0.09	-7.07	-0.11	0.31
Difference			8.59				0.25
Elementary school students	593.12	598.50	-5.38	-0.08	-6.59	-0.10	0.21
Middle/high school students	681.34	668.06	13.28	0.19	18.81	0.27	0.07
Difference			-18.66*				0.03
Reading performance below median	592.88	592.49	0.39	0.01	0.51	0.01	0.94
Reading performance above median	638.83	639.27	-0.44	-0.01	-0.55	-0.01	0.93
Difference			0.83				0.91
Mathematics performance below median	602.27	595.63	6.64	0.09	8.44	0.12	0.25
Mathematics performance above median	633.48	638.11	-4.63	-0.06	-5.92	-0.08	0.34
Difference			11.27				0.13

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. *TerraNova Third Edition* reading and mathematics tests administered three years after application.

Table C-3. Impact estimates of the offer and use of a scholarship on chronic absenteeism (percentage of students absent 10 percent or more days during the school year) after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	21.9	27.3	-5.4*	-0.12	-7.5	-0.17	0.03
Subgroups							
SINI	24.7	30.4	-5.7	-0.12	-8.1	-0.18	0.06
Not SINI	14.9	19.8	-4.8	-0.12	-6.3	-0.16	0.28
Difference			-0.9				0.86
Elementary school students	18.0	18.5	-0.5	-0.01	-0.6	-0.02	0.88
Middle/high school students	28.9	44.0	-15.2*	-0.30	-22.8	-0.46	<0.01
Difference			14.7*				0.01
Reading performance below median	22.6	32.6	-10.1*	-0.21	-14.0	-0.30	0.01
Reading performance above median	21.3	21.8	-0.6	-0.01	-0.8	-0.02	0.87
Difference			-9.5				0.07
Mathematics performance below median	26.9	31.1	-4.2	-0.09	-5.9	-0.13	0.27
Mathematics performance above median	16.9	23.6	-6.7	-0.16	-9.1	-0.21	0.05
Difference			2.5				0.62

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. Student attendance records from the Office of State Superintendent for Education and from private schools for school years 2014-15, 2015-16, and 2016–17.

Table C-4. Impact estimates of the offer and use of a scholarship on parent satisfaction after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	81.6	80.5	1.0	0.03	1.3	0.03	0.72
Subgroups							
SINI	79.5	78.6	0.9	0.02	1.1	0.03	0.80
Not SINI	87.9	86.5	1.4	0.04	1.7	0.05	0.74
Difference			-0.5				0.93
Elementary school students	80.6	81.6	-1.0	-0.03	-1.2	-0.03	0.78
Middle/high school students	83.1	78.2	4.9	0.12	6.8	0.16	0.27
Difference			-6.0				0.29
Reading performance below median	81.4	76.4	5.0	0.12	6.4	0.15	0.84
Reading performance above median	85.1	83.9	1.2	0.03	1.5	0.04	0.73
Difference			-0.4				0.95
Mathematics performance below median	78.0	77.2	0.8	0.02	1.0	0.02	0.85
Mathematics performance above median	85.0	83.8	1.2	0.03	1.6	0.04	0.74
Difference			-0.4				0.94

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. Parent surveys for OSP evaluation, 2015–2017.

Table C-5. Impact estimates of the offer and use of a scholarship on student satisfaction after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	68.5	60.1	8.4*	0.17	11.0	0.22	0.04
Subgroups							
SINI	68.0	60.3	7.8	0.16	10.1	0.21	0.09
Not SINI	70.6	58.3	12.4	0.25	17.7	0.36	0.18
Difference			-4.6				0.66
Elementary school students	74.3	67.1	7.2	0.15	9.0	0.19	0.15
Middle/high school students	60.4	50.5	9.9	0.20	13.9	0.28	0.15
Difference			-2.7				0.76
Reading performance below median	65.6	60.6	5.0	0.10	6.7	0.14	0.41
Reading performance above median	70.8	59.6	11.3*	0.23	14.4	0.29	0.04
Difference			-6.3				0.45
Mathematics performance below median	64.9	58.9	6.0	0.12	7.6	0.15	0.33
Mathematics performance above median	72.6	62.0	10.6	0.22	14.2	0.29	0.05
Difference			-4.7				0.57

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. Student surveys for OSP evaluation, 2015–2017.

Table C-6. Impact estimates of the offer and use of a scholarship on parent perceptions that school is very safe after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	65.9	62.1	3.8	0.08	4.8	0.10	0.27
Subgroups							
SINI	62.5	60.3	2.1	0.04	2.7	0.06	0.63
Not SINI	73.8	66.4	7.4	0.15	9.1	0.19	0.19
Difference			-5.3				0.47
Elementary school students	67.9	67.5	0.4	0.01	0.5	0.01	0.92
Middle/high school students	61.3	51.0	10.3	0.21	14.3	0.29	0.10
Difference			-9.9				0.18
Reading performance below median	62.0	61.6	0.4	0.01	0.6	0.01	0.93
Reading performance above median	69.3	62.0	7.2	0.15	9.0	0.19	0.13
Difference			-6.8				0.31
Mathematics performance below median	62.5	65.3	-2.8	-0.06	-3.5	-0.07	0.59
Mathematics performance above median	69.7	59.6	10.1*	0.20	12.7	0.26	0.03
Difference			-12.9				0.06

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. Parent surveys for OSP evaluation, 2015–2017.

Table C-7. Impact estimates of the offer and use of a scholarship on student perceptions that school is very safe after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	60.5	48.7	11.8*	0.24	16.8	0.34	0.01
Subgroups							
SINI	60.8	48.1	12.7*	0.25	18.0	0.36	0.01
Not SINI	59.4	53.3	6.1	0.12	9.4	0.19	0.60
Difference			6.7				0.60
Elementary school students	59.8	52.9	6.8	0.14	9.3	0.19	0.24
Middle/high school students	60.2	42.4	17.9*	0.36	27.1	0.54	0.01
Difference			-11.0				0.20
Reading performance below median	64.2	50.1	14.2*	0.28	20.6	0.41	0.03
Reading performance above median	55.4	46.6	8.8	0.18	12.3	0.25	0.15
Difference			5.3				0.53
Mathematics performance below median	59.9	46.3	13.5*	0.27	19.8	0.40	0.04
Mathematics performance above median	61.4	51.2	10.2	0.20	14.3	0.29	0.09
Difference			3.3				0.70

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. Student surveys for OSP evaluation, 2015–2017.

Table C-8. Impact estimates of the offer and use of a scholarship on parent involvement in school after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	22.2	22.3	-0.06	-0.01	-0.08	-0.01	0.92
Subgroups							
SINI	21.4	21.4	0.02	<0.01	0.02	0.00	0.98
Not SINI	24.6	24.9	-0.25	-0.03	-0.30	-0.03	0.81
Difference			0.26				0.85
Elementary school students	23.3	25.0	-1.76*	-0.20	-2.12	-0.24	0.02
Middle/high school students	20.1	17.0	3.11*	0.38	4.31	0.52	<0.01
Difference			-4.87*				<0.01
Reading performance below median	21.2	22.3	-1.09	-0.11	-1.38	-0.14	0.26
Reading performance above median	22.9	21.9	0.97	0.11	1.21	0.13	0.24
Difference			-2.06				0.11
Mathematics performance below median	21.6	22.8	-1.15	-0.11	-1.46	-0.14	0.20
Mathematics performance above median	22.8	21.8	0.97	0.11	1.21	0.14	0.24
Difference			-2.12				0.08

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. Parent surveys for OSP evaluation, 2015–2017.

Table C-9. Impact estimates of the offer and use of a scholarship on parent involvement at home after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	17.9	18.2	-0.32	-0.04	-0.40	-0.05	0.49
Subgroups							
SINI	16.8	17.1	-0.32	-0.04	-0.40	-0.05	0.60
Not SINI	20.2	20.6	-0.33	-0.05	-0.40	-0.05	0.67
Difference			0.02				0.99
Elementary school students	20.1	21.2	-1.14*	-0.17	-1.36	-0.21	0.04
Middle/high school students	13.5	12.3	1.23	0.17	1.72	0.23	0.15
Difference			-2.36*				0.02
Reading performance below median	17.8	17.8	-0.02	<0.01	-0.03	<0.01	0.97
Reading performance above median	17.9	18.5	-0.56	-0.07	-0.69	-0.08	0.43
Difference			0.54				0.57
Mathematics performance below median	18.3	18.7	-0.37	-0.05	-0.47	-0.06	0.55
Mathematics performance above median	17.4	17.7	-0.27	-0.03	-0.34	-0.04	0.70
Difference			-0.10				0.91

*Difference between elementary and secondary groups was statistically significant at the 0.05 level.

NOTE: ITT refers to the intent-to-treat impact estimates. TOT refers to the treatment-on-treated impact estimates.

SOURCE: Estimated means and impacts were generated from the study's regression models, as described in appendix section B-3. Parent surveys for OSP evaluation, 2015–2017.

This page intentionally left blank.

Appendix D. Supporting Analyses

This appendix provides supplementary analyses that support reported findings. The first two sections (D-1 and D-2) investigate the possibility that sample attrition (students leaving the study) influenced study impacts, and might help explain the variation of impacts across years for the student achievement and student survey outcomes. The next three sections (D-3, D-4, and D-5) explore additional findings related to the student achievement impacts, describing achievement patterns over time for students who were tested in all three followup years, examining results for treatment group students who used or did not use the scholarship to see if the two groups differ, and presenting additional information about student mobility across schools. Sections D-6 and D-7 report results for alternate measures of student absenteeism and the SINI subgroup. The final section (D-8) presents response frequencies for individual survey items related to parent satisfaction, student safety, and parent involvement as a supplement to the main parent and student outcomes that are presented in the report.

D-1. Analyses of Attrition Related to Achievement Impacts

Attrition in the study sample is a concern because it can lead to bias in measuring impacts, particularly if there is a substantial difference in attrition rates between the treatment and control groups. The analyses in this section show that students tested in the third year are generally similar to the original sample. They also show that while there are some differences between the treatment and control groups tested in the third year, these differences do not appear to be related to baseline achievement.

To examine whether students tested in the third year are systematically different from the full study sample, we first compared their baseline characteristics to see if there were any significant differences (table D-1). The comparison did not show many differences, and differences are generally small. One way of looking at the magnitude of differences is to calculate standardized differences (differences divided by the standard deviation of the characteristic). The standardized differences were all below 0.10, which falls within the acceptable range established by the What Works Clearinghouse, and none were statistically significant.[14]

Next we examined whether nonresponse affected the treatment and control groups differently, which could contribute to the pattern of observed achievement impacts. A lower proportion of the control group completed tests in the third year (62 percent) compared to the treatment group (73 percent). Whether this lower rate contributed to the observed impact depends on whether control group students who were tested differ systematically from treatment group students who were tested. The average test scores for treatment and control students at the time of application were similar (differences of 3 scale score points in reading and 1 scale score point in mathematics, neither of which were significant). Because a range of factors could be related to attrition, the study also estimated a logistic model of response that included the same baseline characteristics that were in the impact models as covariates, but

[14] The What Works Clearinghouse uses standardized differences to assess how different the treatment and control groups are. These differences are calculated as an effect size (difference between the treatment group average and the control group average, divided by a measure of how much the value of the characteristic varies across students or parents). Larger values indicate that samples are more different.

with an outcome for "tested" and "not tested." The results suggested that receiving a scholarship offer was correlated with a greater likelihood of students completing a test (p-value<0.0001). However, test scores at the time of application were not significant predictors of nonresponse ($p = 0.30$ for mathematics and $p = 0.72$ for reading). The results suggest that after controlling for baseline characteristics, there is still significant differential attrition but baseline achievement levels do not appear to be related to that attrition.

Table D-1. Characteristics of original sample and responder sample for student reading test three years after application

	Original sample (base weights)		Responder sample (nonresponse weights)		
Characteristic	Mean	Standard deviation	Mean	Standard deviation	Standardized difference
Entering grade					
Kindergarten	25.5%	43.6%	27.6%	44.7%	0.05
Grade 1	11.4%	31.7%	11.9%	32.3%	0.02
Grade 2	9.7%	29.7%	10.0%	30.0%	0.01
Grade 3	9.3%	29.0%	9.0%	28.7%	0.01
Grade 4	8.4%	27.7%	8.3%	27.7%	0.00
Grade 5	6.0%	23.8%	6.1%	23.9%	0.00
Grade 6	8.3%	27.6%	8.7%	28.1%	0.01
Grade 7	6.0%	23.7%	4.4%	20.6%	0.07
Grade 8	4.7%	21.3%	4.3%	20.2%	0.02
Grade 9	7.0%	25.5%	7.5%	26.3%	0.02
Grade 10	3.7%	18.8%	2.6%	15.9%	0.06
Test score					
Reading scale score at time of application	558.7	92.1	556.3	92.2	0.03
Mathematics scale score at time of application	533.8	111.9	531.9	113.5	0.02
Student characteristics					
Student is female	49.2%	50.0%	50.0%	50.0%	0.02
Student is African American	85.2%	35.5%	85.2%	35.5%	0.00
Student has disabilities or other challenges	13.4%	34.0%	11.0%	31.3%	0.07
Student attends a school in need of improvement	63.2%	48.2%	62.2%	48.5%	0.02
Student age difference from median age of grade	<0.1	0.5	<-0.1	0.4	0.05
Family characteristics					
Parent went to college	59.3%	49.1%	58.5%	49.2%	0.02
Parent gave school grade of A or B at time of application	58.2%	49.3%	59.0%	49.5%	0.02
Parent perception of school safety at time of application	72.5%	44.7%	72.5%	44.9%	0.00
Parent was employed at time of application	46.9%	49.9%	47.9%	50.0%	0.02
Family income in thousands at time of application	12.8	13.5	13.0	13.4	0.02
Number of children in household at time of application	2.6	1.4	2.6	1.4	0.02
Months at current address at time of application (in tens)	6.5	7.8	6.5	7.9	0.01

NOTE: Sample size for original sample in the third year was 1,725 students and responder sample was 1,182 students.

Another approach to exploring whether missing data affected findings is to consider the possible sources of missing data. Of those students not tested in the third year, about two-thirds were not located in a public or private school in DC and the remainder were located in a school but did not complete testing.[15] As discussed below, the available data suggest that the majority of the students not located for testing had moved out of DC, although some may have dropped out.[16] If data are missing for essentially random reasons, there is less evidence of bias. For example, if a student was not tested because a parent had changed jobs and the family moved out of DC, there is no reason to think that student nonresponse is biasing impacts. A concern would arise if data are missing for reasons that cannot be considered random. For example, if families move out of DC in search of better schools elsewhere, whether they were awarded a scholarship may be a factor in their decision.

Unfortunately, the study does not have data on reasons for family moves. However, we did investigate the hypothesis that higher-achieving students in the control group (those who did not receive a scholarship offer) were more likely to move out of DC than higher-performing students in the treatment group (those who received a scholarship offer), which would bias achievement impacts. If treatment and control students who moved out had similar ability levels to those who did not, it is less plausible that impacts were biased (though it would reduce the study's sample sizes and its statistical precision). To test the hypothesis, the study used data from the DC Office of the State Superintendent for Education (OSSE) and participating private schools to identify students who potentially moved or dropped out. The data sources were attendance and enrollment data from OSSE, which included all public-school students in DC (including charter schools), and the scholarship payment data from the scholarship program operator, which indicated if students were attending private schools. The study classified students as "potentially moved or dropped out" if they were not found in any of these data sources. The study could not rule out the possibility that some of these students still lived in DC but were being homeschooled or were attending a non-participating private school.

Of the initial study sample, 16 percent of students potentially moved or dropped out at some time in the previous three years, and students in the treatment group moved or dropped out at a lower rate than the control group. For the treatment group, the rate was 14 percent and for the control group it was 20 percent, a statistically significant difference ($p < 0.001$). As noted above, these differences in rates of "potentially moved or dropped out" would affect impacts if higher-performing students in the control group were more likely than higher-performing students in the treatment group to have potentially moved or dropped out. A comparison of baseline mathematics scores for treatment and control group students who potentially moved or dropped out showed that the two groups had similar scores (553 and 559, respectively) and the difference was not statistically significant ($p = 0.69$).

There were two statistically significant differences for other baseline characteristics. More of the treatment students who potentially moved or dropped out were from SINI schools and were entering eighth grade when they applied. The study's regression models include these characteristics in order to

[15] Refusals were one reason that students were not tested: of those not tested, 5 percent of students in the treatment group and 7 percent of students in the control group refused to take a test.

[16] Students in DC cannot drop out legally until they reach age 18, which would suggest that any student still living in DC should be in the OSSE attendance data file or in private schools. However, the study's approach could not distinguish between students who moved out of DC and students who were living in DC but not attending a traditional public or charter school. Both kinds of students will be missing from the data files. The study uses the expression "moved or dropped out" to acknowledge this.

adjust for such differences. Also, these factors contribute much less to test scores in the third year than baseline test scores contribute. For example, dropping these two characteristics from the mathematics impact model reduced the model's predictive ability by less than 1 percent.[17] Dropping test scores from the model reduced the model's predictive ability by 21 percent.

In summary, these analyses show that control group students were more likely than treatment group students to potentially move or drop out, but their ability levels at the time of application were not statistically different from treatment group students who potentially moved or dropped out. It does not appear that the higher rate of moved or dropped out introduced systematic bias to the analyses.

D-2. Analyses of Attrition Related to Outcomes from Student Survey

This section examines to what extent the impacts on student satisfaction and student perceptions might be affected by the response rates on the student survey. Three years after applying, the study found that the OSP program had a positive impact on student satisfaction and student perceptions of school safety. The student survey had a relatively low response rate (62 percent) and the response rate differed between the treatment and control groups by 11 percentage points.[18] The low overall rate and the differential between groups potentially leads to an incorrect measure of the program's impact, according to the What Works Clearinghouse attrition standard. The incorrectness would arise through some combination of students in the control group who did not respond to the survey being more likely to report satisfaction with their school and/or schools being safer, and students in the treatment group who did not respond to the survey being less likely to report these perceptions.

Due to this concern, we assessed whether the impacts based on the student survey potentially were affected by nonresponse by estimating a model in which whether students responded was a function of covariates used in the impact models. There was more reason to be concerned about nonresponse if it was correlated with other variables. (If nonresponse was random, it acted the same as shrinking the sample size without affecting other aspects of the groups). The results indicated that response was correlated with the treatment indicator and three of the 17 covariates, family income, the difference between a student's age and the median age of the grade level (the study's variable denoting whether students were over age for their grade), and whether a student was female (table D-2). Students were more likely to respond if they were in the treatment group, had higher family income, or were female, and less likely to respond if they were over age for their grade.

[17] The marginal R^2 (explained variance) was reduced from 0.472 to 0.471. Dropping the baseline test scores from the model reduces the marginal R^2 from 0.472 to 0.372.

[18] The number of students eligible to complete the survey was larger than in previous years because the survey's grade range stayed the same and more students aged into the range than aged out of it. Students who were entering second and third grade when they first applied for the scholarship were in fourth or fifth grade in the third year after application and were given the survey. Students who were entering eleventh or twelfth grade at the time of application were no longer in the survey because they had exited twelfth grade by the time of the third followup, but the number of older students no longer surveyed (46) was outweighed by the number of younger students who were added to the survey (330).

Table D-2. Significant coefficients from model of response to student survey

Variable	Coefficient	*p*-value
Treatment status	0.134	<0.0001
Family income (in $1,000s)	0.002	0.0479
Difference from median age	-0.064	0.0230
Student is female	0.085	0.0062

SOURCE: Coefficients were generated from the study's regression models, as described in appendix section B-3. Student surveys for OSP evaluation, 2015–2017.

These significant correlations suggest that impacts *could* be mismeasured, but are not evidence that they *were* mismeasured. To explore the issue further, we introduced a possibility that both nonresponse and the safety outcome were correlated with a variable that was not observed, termed a "hidden variable" in the literature (see Rosenbaum and Rubin 1983; Imbens and Rubin 2015, chapter 22). As Imbens and Rubin note, in most research contexts, failing to account for this hidden variable is likely to have a smaller impact on findings than failing to account for the variables that are not hidden. Studies typically collect data on variables deemed most likely to be correlated with outcomes.

To operationalize this insight, thirteen regression models were run in which the impact on student safety was measured leaving out one covariate at a time (each covariate became a hidden variable). The results suggest the impact reported in the main text is unlikely to be the result of a hidden variable (tables D-3 and D-4). The result with all covariates in the model was an impact of 8.5 percentage points for student satisfaction and 11.5 percentage points for student perception of school safety. Estimates ranged between 7.9 percent and 8.8 percent for student satisfaction, and between 11.3 percent and 12.2 percent for student perceptions of school safety. These estimates remained close to the impacts for the full model.

This analysis does not mean there was no hidden variable. It indicates that the impact measure was robust to 14 different covariates being one of the hidden variables. For a truly hidden variable to affect results more, it would need to be both correlated with nonresponse *and* correlated with the outcome to a stronger degree than any of the 14 variables examined here. Considering the range covered by these variables, it is difficult to think what that variable could be.

Table D-3. Sensitivity of student satisfaction impact estimate to dropping covariates

	Impact estimate	p-value
Full Model	8.5%	0.039
Covariate dropped		
Reading score	8.5	0.040
Mathematics score	8.8	0.034
Student is female	8.5	0.040
Student is black	8.4	0.042
Student has disability or other challenges	8.6	0.036
Student attended a SINI school	8.4	0.042
Student age difference from median age of grade	7.9	0.056
Parent has any college education	8.5	0.038
Parent rating of school satisfaction	8.4	0.041
Parent rating of school safety	8.4	0.042
Parent is employed	8.3	0.044
Household income	8.0	0.056
Number of children in household	8.7	0.036
Months at current address	8.2	0.046

NOTE: All covariates were measured at the time of application.

SOURCE: Estimates were generated from the study's regression models, as described in appendix section B-3. Student surveys for OSP evaluation, 2015–2017.

Table D-4. Sensitivity of student safety impact estimate to dropping covariates

	Impact estimate	p-value
Full Model	11.5%	0.011
Covariate dropped		
Reading score	12.1	0.008
Mathematics score	12.2	0.007
Student is female	11.8	0.009
Student is black	11.9	0.009
Student has disability or other challenges	11.8	0.009
Student attended a SINI school	11.8	0.009
Student age difference from median age of grade	11.4	0.012
Parent has any college education	11.7	0.010
Parent rating of school satisfaction	11.8	0.009
Parent rating of school safety	11.8	0.009
Parent is employed	11.6	0.010
Household income	11.3	0.013
Number of children in household	11.4	0.012
Months at current address	11.9	0.008

NOTE: All covariates were measured at the time of application.

SOURCE: Estimates were generated from the study's regression models, as described in appendix section B-3. Student surveys for OSP evaluation, 2015–2017.

D-3. Comparing Impacts Between the Study's Followup Years

While the previous two reports found that OSP had negative effects on mathematics achievement one and two years after applying to the program, the current report found no significant impact on this outcome after three years. This pattern seems to suggest that the typical scholarship recipient had lower mathematics achievement in the first two years than they would otherwise have had—but caught up by the third year. This section tests and ultimately rejects two alternative explanations for the pattern of impacts over time. The analysis described in this section provides support for the interpretation that scholarship recipients caught up with their peers in the third year.

Two alternative explanations for the pattern of findings across the three years are described below:

1. **Differences in the students tested each year.** Not every student in the sample was tested each year. This raises the question of whether the pattern of impacts could be due to differences in the study sample across the three years.

2. **Random error in the impact estimates.** Even if the true impact was identical in all three years, some variation in the impact estimates across the years would be expected due to imprecision in the estimates.

To test the first alternative explanation, impacts in each year were re-estimated for the longitudinal sample of students who were tested in all three years. With a longitudinal sample, differences in impacts across years cannot be due to differences in the students tested. The findings are consistent with the pattern described earlier—negative impacts in the first two years and no significant impact in the third year (table D-5). This indicates that differences in the students tested across years were not responsible for the observed pattern of impact estimates over time.

Table D-5. Impact estimates and adjusted means of the offer of a scholarship on reading and mathematics scale scores, by year for the longitudinal sample

	Reading scale score			Mathematics scale score		
Year	Treatment group mean	Control group mean	Difference (estimated impact)	Treatment group mean	Control group mean	Difference (estimated impact)
Baseline	567.37	566.49	0.88	537.80	538.00	-0.20
Year 1	601.09	604.51	-3.42	578.54	582.98	-4.44
Year 2	618.96	623.58	-4.62	595.06	606.47	-11.41*
Year 3	632.94	633.22	-0.29	619.87	619.61	0.26

*Difference between the treatment group and the control group was statistically significant at the 0.05 level. The mathematics impacts for Year 2 and Year 3 were also significantly different at the 0.05 level.

NOTE: Sample size was 676 students for reading and 672 students for mathematics. Impacts reported here for the longitudinal sample (i.e., students tested at baseline and in all three followup years) differ from previously reported estimates for the impact-year samples.

SOURCE: Estimated means and impacts for the longitudinal sample were generated from the study's regression models, as described in appendix section B-3. The treatment and control means for each year are regression-adjusted to account for baseline differences and evaluated at the sample mean across both groups. *TerraNova Third Edition* reading and mathematics tests.

To test the second alternative explanation, the study conducted a formal test of whether the impact on mathematics achievement was the same in all three years. If so, it would suggest that random error is a likely explanation for the transition from negative to zero impact estimates on mathematics achievement in the third year. Put differently, it would cast doubt on the conclusion that the impact on mathematics achievement improved in the third year as scholarship recipients caught up to their peers.

To conduct this test, the study applied multivariate regression analysis to the longitudinal sample of students tested in all three years.[19] The test results revealed statistically significant variation in impacts across the three years ($p = 0.02$). In addition, the difference in impacts between the second and third years was statistically significant ($p = 0.03$). This suggests that the true impact of the OSP on mathematics achievement varied across the three years and that the trend from negative to zero impact between the second and third years was not due to random error in the impact estimates.

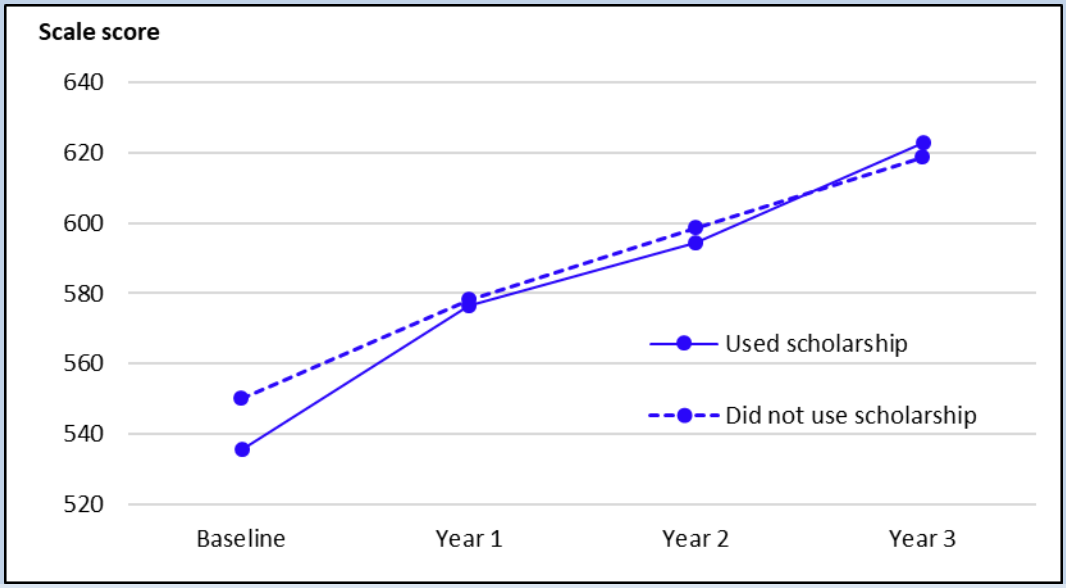
D-4. Analysis of Achievement Differences Within Treatment Group

To better understand differences between the study's second-year and third-year impact on mathematics test scores, we examined how average mathematics test scores of the treatment group changed over time, separately for students who used the scholarship or did not use it. This analysis used the longitudinal sample of students who were tested in all years, which enabled us to examine test scores for the group of students whose composition did not change over time.

Figure D-1 shows mathematics test scores for the two groups of students: 1) treatment group students who used the scholarship to attend a private school at any point during the three years and 2) treatment group students who did not use the scholarship and attended public schools in all three years. Treatment group students who did not attend private schools showed similar gains in test scores compared with treatment group students who attended private schools. Looking separately at students who used their scholarships and students who did not would not have altered the results.

[19] This analysis accounted for the statistical dependence or correlation in the impact estimates across the three years since all three impacts were estimated using the same sample of students.

Figure D-1. Average mathematics test scores for treatment group students in the longitudinal sample, by scholarship use and year



Group	Baseline	Year 1	Year 2	Year 3
Used scholarship offer to attend private school	535.45	576.43	594.46	622.70
Did not use scholarship offer, attended public school	550.05	578.26	598.61	618.65

NOTE: Average scale scores were calculated based on the longitudinal sample of 423 treatment group students who were tested in mathematics at baseline and all three followup years (347 who used the scholarship in Year 3 and 76 who did not use the scholarship in Year 3). Baseline scores for the two groups were significantly different at the 0.05 level. Differences for Year 1, Year 2, and Year 3 were not statistically significant.

SOURCE: Estimated means were generated from the study's regression models, as described in appendix section B-3. The treatment and control means for each year are regression-adjusted to account for baseline differences and evaluated at the sample mean across both groups. *TerraNova Third Edition* mathematics test.

D-5. Measuring the Mobility of Students Across Schools

The study examined student mobility as a possible explanation for the changes in mathematics achievement impacts in the third year, but found that the mobility rate between the second and third years was comparable for students in the treatment and control groups. The study also examined the number of times students changed schools over the three years to see whether there were differences between the two groups. Control group students were less likely to change schools more than once (29 percent) compared with students in the treatment group (36 percent) (table D-6). This analysis further supports the finding that there is little evidence that differences in student mobility across schools for the treatment and control groups explains the changes in mathematics achievement impacts in the third year.

Table D-6. Percentage of students by the total number of school changes, by group

School change	Treatment group	Control group
Did not change schools	5	16
Changed schools once	59	55
Changed schools twice	29	25
Changed schools three times	7	5

NOTE: Percents may not sum to 100 because of rounding. The difference between the treatment and control groups in the total number of school changes was significant (p<.0001).

D-6. Measuring Student Absenteeism

In addition to measuring the impact of the scholarship on student chronic absenteeism, the study also measured the impact on percentage of days absent. Students who received a scholarship offer were absent on a lower percentage of school days by 1.9 percentage points (table D-7), 7.1 percent compared with 9.1 percent.

Table D-7. Impact estimates of the offer and use of a scholarship on the percentage of school days absent after three years

	Impact of scholarship offer (ITT)				Impact of scholarship use (TOT)		
Sample	Treatment group mean percentage	Control group mean percentage	Difference (estimated impact)	Effect size	Adjusted impact estimate	Effect size	*p*-value of estimates
Full sample	7.1	9.1	-1.9	-0.18	-2.7	-0.25	<0.01

SOURCE: Estimates were generated from the study's regression models, as described in appendix section B-3. Student attendance records from the Office of State Superintendent for Education and from private schools for school years 2014-15, 2015-16, and 2016–17.

D-7. Impacts on Test Scores in SINI and Non-SINI Schools, Excluding Pre-K Students

This section examines whether test score impacts for the group of students attending schools in need of improvement at the time of application might be affected by students who were entering pre-K when they applied. Students in grades K–12 are eligible for OSP scholarships, which means students can be attending pre-K programs at the time their parents apply for a scholarship. In fact, nearly a quarter of the study sample was attending pre-K. Because the legislation required that the lottery give priority to students from SINI schools, the program needed to categorize students as attending SINI schools or not, and pre-K students were all categorized as attending non-SINI schools even though some of them might be attending a public school that had been designated as SINI. Preschool programs do not fall within statutory definitions of SINI. One implication is that this categorization combines pre-K students with older students in grades K–12 who are attending higher-performing schools.

Results for mathematics test scores showed a positive impact for SINI students compared with a negative impact for non-SINI students, although neither impact was statistically significant. To assess if the results were related to the categorizing of all pre-K as non-SINI, test-score models were estimated with pre-K students excluded from the sample. Excluding pre-K students yielded larger negative impacts in mathematics for non-SINI students, while the impacts for SINI students changed only slightly (table D-8).

Table D-8. Comparing subgroup impacts with and without pre-K students, after three years

	Reading				Mathematics			
	SINI		Non-SINI		SINI		Non-SINI	
	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Including pre-K	-2.19	0.44	-0.83	0.85	2.91	0.55	-5.68	0.31
Excluding pre-K	-2.40	0.40	-1.01	0.90	2.68	0.58	-14.33	0.12

SOURCE: Estimates were generated from the study's regression models, as described in appendix section B-3.

D-8. Supplemental Tables for Parent and Student Survey Items

This section presents results for individual survey items on the parent and student surveys that asked about parent satisfaction, student safety, and parent involvement three years after application.

Parent Satisfaction

In addition to rating their child's school with a letter grade as the main measure of satisfaction, parents also provided ratings of their satisfaction with 16 specific aspects of their child's school. Simple comparisons of the percentage of parents who chose one of four responses—which corresponded to very dissatisfied, dissatisfied, satisfied, and very satisfied—were informative about what may be driving the letter grades that parents gave schools. Ten of the 16 items were significantly higher for the treatment group (table D-9). For example, 44 percent of treatment group parents were "very satisfied" with academic quality compared with 35 percent of control group parents.

Student Safety

In addition to a question about overall school safety, which was the main outcome analyzed in the text, the student survey also asked whether various negative events had happened to students at school. Students indicated whether the events had happened to them never, once or twice, or three or more times. Treatment and control group proportions for each of the eight items are shown in table D-10. There were two statistically significant differences between the treatment and control group. Students in the control group were more likely to report having been threatened with physical harm at school or having been bullied at school.

Table D-9. Percentage of parents reporting satisfaction with specific aspects of their child's school three years after application

How satisfied are you with the following aspects of this child's current school?	Treatment	Control	*p*-value
Location of school			0.25
Very dissatisfied	2.3	3.0	
Dissatisfied	7.8	9.8	
Satisfied	44.4	46.9	
Very satisfied	45.6	40.3	
School safety			<0.01*
Very dissatisfied	1.9	3.9	
Dissatisfied	7.3	8.2	
Satisfied	44.9	53.4	
Very satisfied	46.0	34.5	
Class sizes			<0.01*
Very dissatisfied	2.6	3.8	
Dissatisfied	7.7	9.6	
Satisfied	47.1	56.6	
Very satisfied	42.5	30.1	
School facilities			0.15
Very dissatisfied	1.5	1.7	
Dissatisfied	8.4	9.3	
Satisfied	53.0	58.5	
Very satisfied	37.2	30.5	
Respect between teachers and students			<0.01*
Very dissatisfied	3.3	4.0	
Dissatisfied	8.8	8.4	
Satisfied	41.6	52.1	
Very satisfied	46.4	35.5	
How much teachers inform parents of students' progress			<0.01*
Very dissatisfied	3.5	1.9	
Dissatisfied	8.6	11.3	
Satisfied	39.3	49.5	
Very satisfied	48.6	37.3	
How much students can observe religious traditions			<0.01*
Very dissatisfied	3.1	9.2	
Dissatisfied	8.6	15.0	
Satisfied	45.6	51.7	
Very satisfied	42.7	24.1	
Academic quality			<0.01*
Very dissatisfied	3.5	3.5	
Dissatisfied	8.9	10.2	
Satisfied	43.4	51.6	
Very satisfied	44.3	34.7	

See notes at end of table.

Table D-9. Percentage of parents reporting satisfaction with specific aspects of their child's school three years after application—Continued

How satisfied are you with the following aspects of this child's current school?	Treatment	Control	*p*-value
Parental involvement in the school			<0.01*
Very dissatisfied	2.6	3.8	
Dissatisfied	7.5	12.0	
Satisfied	51.3	56.2	
Very satisfied	38.7	28.0	
Discipline at the school			<0.01*
Very dissatisfied	3.7	6.0	
Dissatisfied	10.9	12.5	
Satisfied	44.2	50.6	
Very satisfied	41.3	30.8	
Racial mix of students			0.09
Very dissatisfied	3.2	3.4	
Dissatisfied	12.6	15.9	
Satisfied	49.9	53.2	
Very satisfied	34.4	27.6	
Services for children with special needs			0.50
Very dissatisfied	5.4	6.3	
Dissatisfied	13.0	14.6	
Satisfied	47.9	49.6	
Very satisfied	33.7	29.5	
Access to information about the school through printed materials or the school website			<0.01*
Very dissatisfied	1.9	3.0	
Dissatisfied	6.4	10.3	
Satisfied	49.2	53.3	
Very satisfied	42.5	33.5	
Services for students who struggle academically			0.27
Very dissatisfied	4.9	6.2	
Dissatisfied	13.8	13.8	
Satisfied	48.9	53.0	
Very satisfied	32.3	27.1	
Availability of computers			<0.01*
Very dissatisfied	3.2	2.1	
Dissatisfied	8.1	10.3	
Satisfied	50.6	57.8	
Very satisfied	38.1	29.8	
Teacher absenteeism			0.12
Very dissatisfied	1.9	2.9	
Dissatisfied	6.5	7.9	
Satisfied	54.3	58.4	
Very satisfied	37.3	30.8	

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: To calculate p-values, for each item a chi-squared test (weighted by the composite weight) was conducted so that the distributions of frequencies were the same for the treatment group and the control group. Because the items were not primary outcomes, the p-values had not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

SOURCE: Parent surveys for OSP evaluation, 2015–2017.

Table D-10. Percentage of students reporting negative safety incidents that occurred at school three years after application

Did the following ever happen to you at school this year?	Treatment	Control	*p*-value
Had something stolen from your desk, locker, or other place			0.72
Never	55.5	52.6	
Once or twice	33.0	35.8	
Three times or more	11.6	11.6	
Been forced by other kids to give them money or my stuff			0.91
Never	92.3	93.0	
Once or twice	5.0	4.8	
Three times or more	2.7	2.2	
Been offered drugs			0.88
Never	93.6	93.8	
Once or more times	4.5	4.8	
Three times or more	1.9	1.4	
Been physically hurt by another student			1.00
Never	74.7	74.7	
Once or twice	18.1	18.0	
Three times or more	7.2	7.4	
Been threatened with physical harm			0.03*
Never	79.5	77.8	
Once or twice	16.0	12.8	
Three times or more	4.5	9.5	
Seen anyone with a real or toy gun or knife at school			0.45
Never	86.2	84.3	
Once or twice	10.6	13.3	
Three times or more	3.2	2.4	
Been bullied at school			0.01*
Never	77.4	68.4	
Once or twice	15.9	18.9	
Three times or more	6.8	12.7	
Been called a bad name			0.63
Never	48.2	48.9	
Once or twice	31.2	28.2	
Three times or more	20.6	22.9	

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: To calculate *p*-values, for each item a chi-squared test (weighted by the composite weight) was conducted so that the distributions of frequencies were the same for the treatment group and the control group. Because the items were not primary outcomes, the p-values had not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

SOURCE: Student surveys for OSP evaluation, 2015–2017.

Parent Involvement in Education

Two sets of items from the parent survey were used to create the main measures of parent involvement for the impact study. For parent involvement in education at school, parents indicated whether various school events happened never, once, 2 or 3 times, or 4 or more times. For each item, the study assigned a value of 0, 1, 2.5, or 5, depending on the parent response, and then added the resulting eight numbers. The resulting sum is a general measure of how many times parents participated in the various activities with the child's school.

For education involvement in the home, parents could indicate they never did the activity or did an activity once, 2 or 3 times, 4 or 5 times, or 6 or more times. The study used the same procedure described to construct a general measure of involvement, by assigning values to each category (in this case, the values were 0, 1, 2.5, 4.5, and 7), and summing the numbers for the four items.

For individual items that made up the general measures, only one of the differences in parent involvement was statistically significant. Parents of student's in the treatment group were more likely to report receiving information about their child's school at a higher frequency (4 or more times) during the school year using means such as newsletters and school notices (table D-11).

Table D-11. Percentage of parents reporting involvement in education activities at school three years after application

During this school year, how often did you do the following related to this child's school…	Treatment	Control	*p*-value
Receive report cards about this child's performance			0.51
Never	1.8	1.5	
Once	3.3	2.7	
2 or 3 times	53.8	50.3	
4 or more times	41.1	45.5	
Receive information about this child's school, such as newsletters and school notices			<0.01*
Never	5.3	3.1	
Once	3.7	6.0	
2 or 3 times	18.4	25.5	
4 or more times	72.6	65.5	
Communicate with a teacher informally (in person, by phone, or via email)			0.05
Never	3.1	4.6	
Once	5.7	7.6	
2 or 3 times	24.7	29.2	
4 or more times	66.5	58.6	
Attend parent-teacher conferences			0.68
Never	7.2	6.2	
Once	12.9	15.2	
2 or 3 times	44.8	43.5	
4 or more times	35.1	35.2	
Attend school activities for families (dinners, student presentations, open houses, family mathematics, or science nights)			0.13
Never	15.5	16.5	
Once	13.6	18.3	
2 or 3 times	37.2	35.6	
4 or more times	33.8	29.6	
Volunteer in the school			0.51
Never	39.2	42.1	
Once	16.2	17.7	
2 or 3 times	25.5	22.1	
4 or more times	19.1	18.1	
Attend a PTA meeting (or other similar organization meeting)			0.19
Never	24.3	26.4	
Once	15.8	18.1	
2 or 3 times	35.8	29.6	
4 or more times	24.2	25.9	
Accompany students on class trips			0.63
Never	56.3	55.4	
Once	15.6	16.0	
2 or 3 times	17.8	16.0	
4 or more times	10.4	12.6	

*Difference between the treatment group and the control group was statistically significant at the 0.05 level.

NOTE: To calculate *p*-values, for each item a chi-squared test (weighted by the composite weight) was conducted so that the distributions of frequencies were the same for the treatment group and the control group. Because the items were not primary outcomes, the p-values had not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.

SOURCE: Parent surveys for OSP evaluation, 2015–2017.

Table D-12. Percentage of parents reporting involvement in education activities at home three years after application

In the past month, how often did you do the following…	Treatment	Control	*p*-value
Help this child with his or her homework			0.58
Never	9.6	9.7	
Once	5.8	6.9	
2 or 3 times	19.2	16.9	
4 or 5 times	12.6	15.3	
6 or more times	52.9	51.2	
Help this child with reading or mathematics that was not part of his or her homework			0.62
Never	16.9	15.2	
Once	4.2	5.7	
2 or 3 times	19.5	21.5	
4 or 5 times	14.2	14.9	
6 or more times	45.2	42.7	
Talk to this child about his or her experiences in school			0.36
Never	1.4	1.8	
Once	1.9	2.9	
2 or 3 times	7.4	9.5	
4 or 5 times	14.2	15.9	
6 or more times	75.1	69.9	
Work with this child on a school project			0.75
Never	16.8	17.8	
Once	16.1	16.9	
2 or 3 times	28.1	24.7	
4 or 5 times	14.3	13.7	
6 or more times	24.8	27.0	

NOTE: To calculate *p*-values, for each item a chi-squared test (weighted by the composite weight) was conducted so that the distributions of frequencies were the same for the treatment group and the control group. Because the items were not primary outcomes, the p-values had not been adjusted for multiple comparisons. Therefore, the statistical significance for individual items should be interpreted with caution.
SOURCE: Parent surveys for OSP evaluation, 2015–2017.

This page intentionally left blank.

References

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434): 444-455.

CTB/McGraw-Hill. (2010). *TerraNova Third Edition Technical Report*. Monterey, CA: Author.

CTB/McGraw-Hill. (2008). *TerraNova, The Third Edition.* Monterey, CA: Author.

Chingos, M. M., and Kuehn, D. (2017). *The Effects of Statewide Private School Choice on College Enrollment and Graduation: Evidence from the Florida Tax Credit Scholarship Program*. Washington, DC: Urban Institute.

Cochran, W. G. (1977). *Sampling Techniques, 3rd Ed*. New York: Wiley.

Dynarski, M., Rui, N., Webber, A., and Gutmann, B. (2017). *Evaluation of the DC Opportunity Scholarship Program: Impacts After One Year (*NCEE 2017-4022). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Feldman, J., Lucas-McLean, J., Gutmann, B., Dynarski, M., and Betts, J. (2015). *Evaluation of the DC Opportunity Scholarship Program: An Early Look at Applicants and Participating Schools Under the SOAR Ac*t (NCEE 2015–4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hungerman, D., Rinz, K., and Frymark, J. (2017). *Beyond the Classroom: The Implications of School Vouchers for Church Finances*. Working Paper No. 23159. Cambridge, MA: National Bureau of Economic Research.

Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences An Introduction.* Cambridge, MA: Cambridge University Press.

Liang, K-Y., and Zeger, S. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika, 73*(1): 13-22.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

Lohr, S. L. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA: Duxbury Press.

Oh, H. L., and F. J. Scheuren (1983). Weighting Adjustment for Unit Nonresponse. In *Incomplete Data in Sample Surveys, Vol 2* (Eds. W. G. Madow, I. Olkin, and D. B. Rubin), 143-184. New York: Academic Press.

Rosenbaum, P. R., and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika, 70*: 41-55.

Potter, F. J. (1990). A Study of Procedures to Identify and Trim Extreme Sampling Weights. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 225-230.

Stocking, M., and Lord, F. M. (1983). Developing a Common Metric in Item Response Theory. *Applied Psychological Measurement, 7*, 207-210.

Valliant, R., Dever, J. A., and F. Kreuter (2013). *Practical Tools for Designing and Weighting Survey Samples*. New York; Springer.

Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., Eissa, N., and Carr, M. (2010). *Evaluation of the DC Opportunity Scholarship Program: Final Report* (NCEE 2010-4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.